Enabling Audiovisual User Interfaces

Duration: 2 years (May/2014 – Apr/2016)
URL: http://avuis.goldsmithsdigital.com

Originality

The two-year project was supported by a Marie Curie EU fellowship, conducted at EAVI research group, Goldsmiths, U. London. It investigated how human-computer interactions can be sonified and visualised in order to improve user experience and usability. To address this issue, a new UI paradigm is proposed – AVUI (AudioVisual User Interface). AVUI links interaction, sound and image, building upon the concept of Graphical User Interface (GUI) by adding interconnected sound and image. Additionally, the author introduced a series of intensive coding events (hackathons) to assist in building the systems, which then led to performances. The main case study relied on audiovisual performance systems.

Significance

The project received funding from the EU Marie Curie Individual Fellowship program (€300k) and was hosted by Goldsmiths. It resulted in 2 performances at Goldsmiths with hackathon participants, and 3 individual performances – in Berlin (in an art gallery, Mono Shop), in London (EAVI Nights) and in Bath (Seeing Sound symposium). It led to 11 invited talks, 6 in the UK (Goldsmiths, Queen Mary U. London, VJ London, London Music Hackspace, MiXD Symposium, Splice Festival), 2 in Russia (Alexandrinsky Theatre), in Poland (Sound Bureau Katowice), in Estonia (Varvara & Mar Studio) and in Portugal (M-ITI). Additionally, it led to 6 peer-reviewed conference papers. Finally, it resulted in a software toolkit released as open-source.

Rigour

A user-centred design methodology was adopted. The project involved multiple stages, over 2 years: a workshop and 3 hackathons (average of 2 days and 15 participants each). An initial workshop gathered ideas from participants. These ideas informed 2 hackathons, where performance systems were created, resulting in a software toolkit. The systems were peer-reviewed by experienced performers. The toolkit was tested in a final hackathon and subsequently released as open-source. These steps involved a mixed-methods approach, combining brainstorming, sketching, interviews and observation.

Table of Contents – Documentation

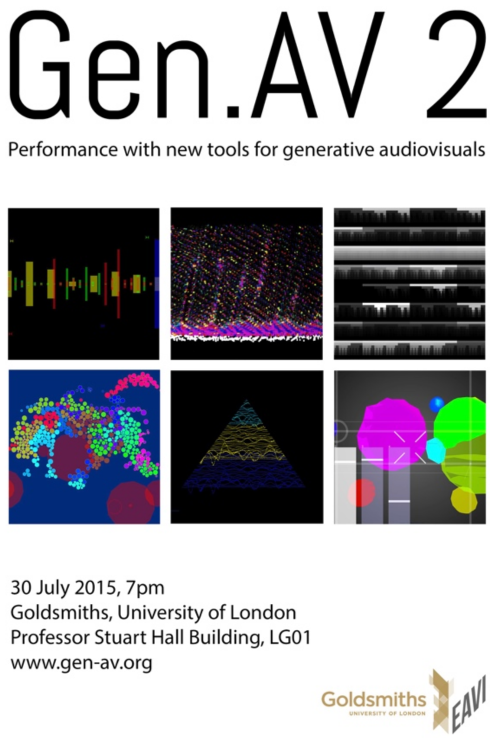
Documentation

Performances

Photos of group performances at Goldsmiths, University of London

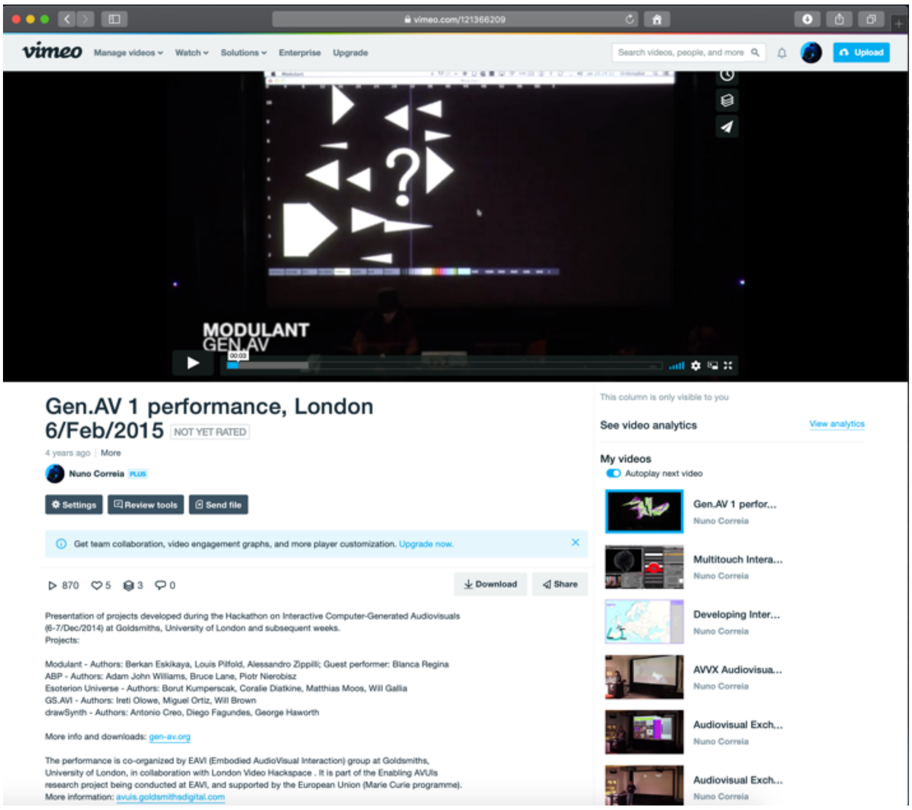
Performance 1, 6/2/2015, (left)
Performance 2, 30/7/2015 (right)

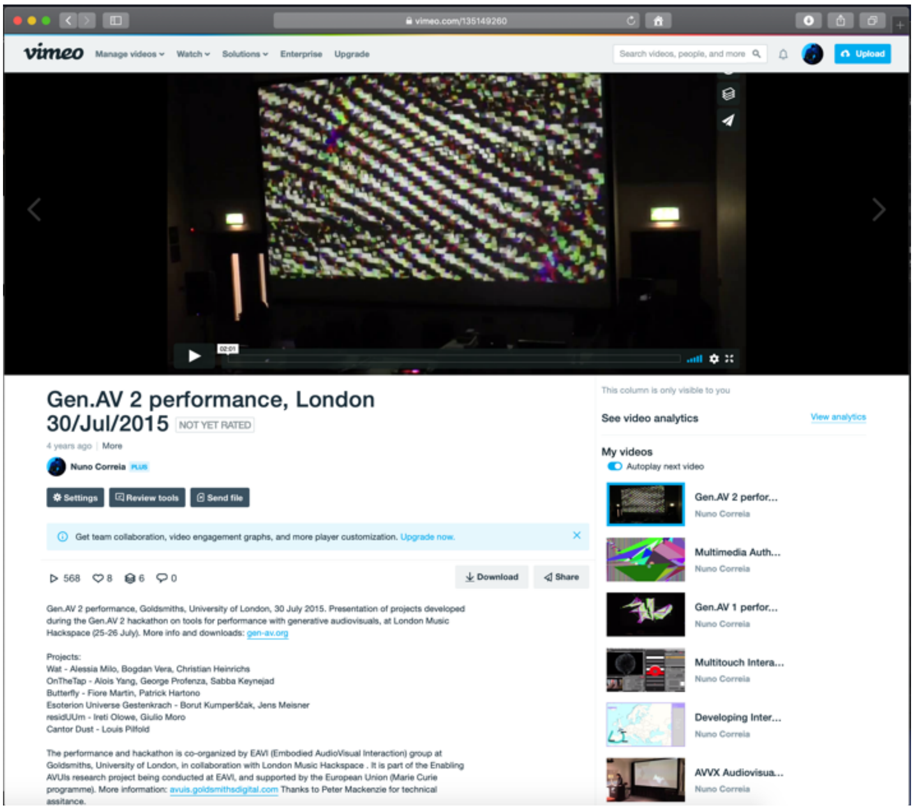
Posters of group performances



Videos of group performances

Group performance 1 https://vimeo.com/121366209
Group performance 2 https://vimeo.com/135149260

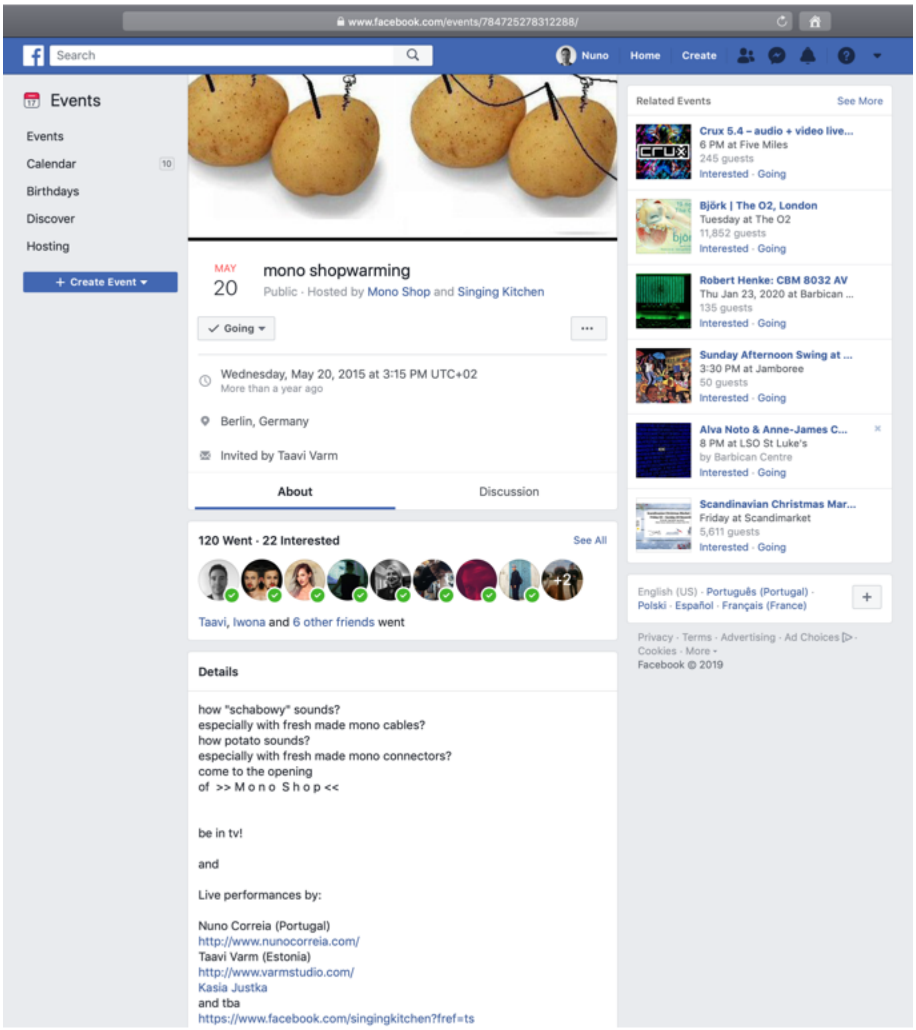


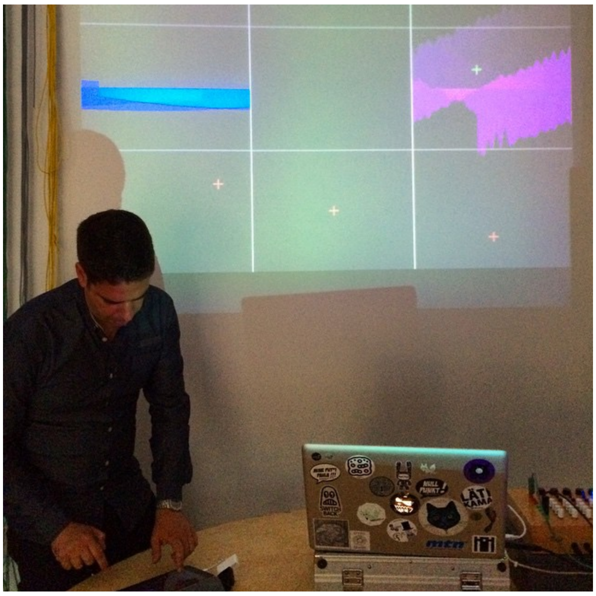


Photos from the workshop and 3 hackathons, preparing group performances

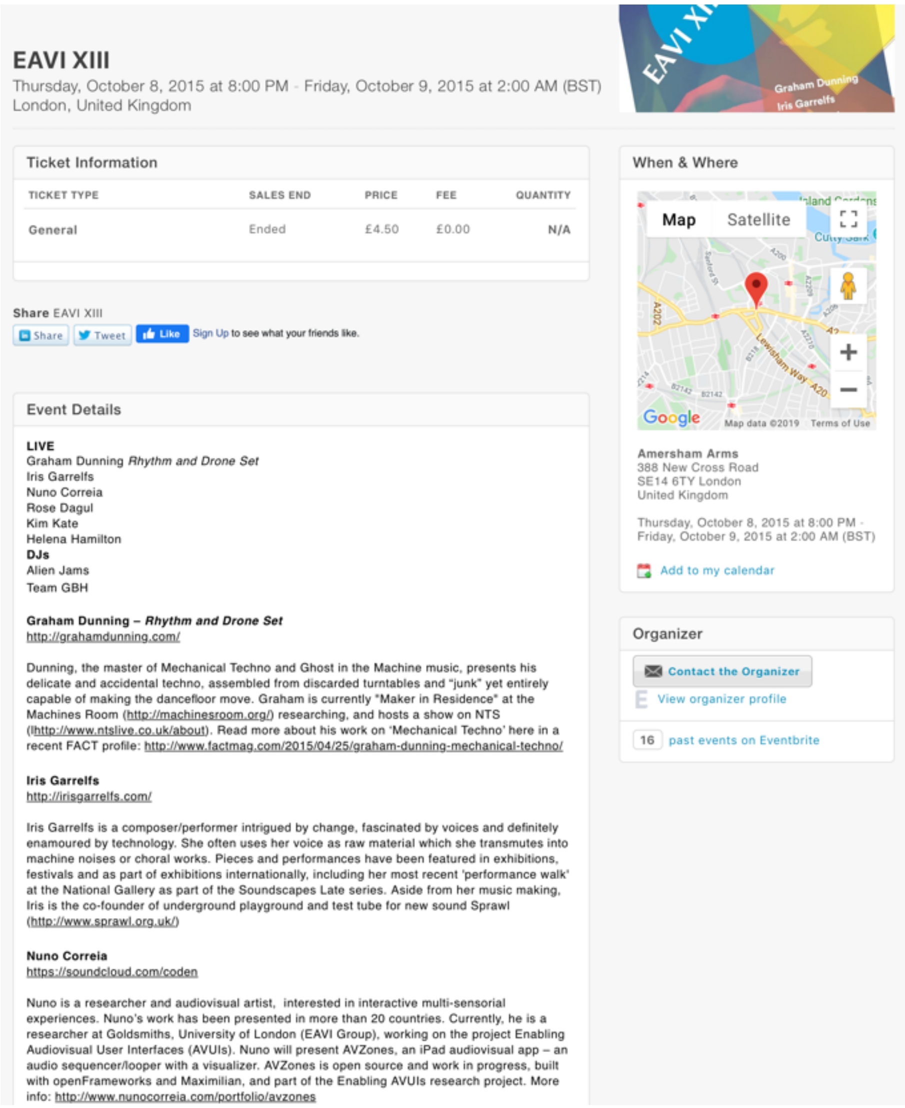



Photos and promotional materials: solo performance, Mono Shop Gallery, Berlin, 20/5/2015





Photos and promotional materials: solo performance, Amersham Arms, 8/10/2015



EAVI XIII

Thursday, October 8, 2015 at 8:00 PM - Friday, October 9, 2015 at 2:00 AM (BST)
London, United Kingdom

Ticket Information

TICKET TYPE	SALES END	PRICE	FEE	QUANTITY
General	Ended	£4.50	£0.00	N/A

Share EAVI XIII

Share Tweet Like Sign Up to see what your friends like.

Event Details

LIVE
Graham Dunning *Rhythm and Drone Set*
Iris Garrelfs
Nuno Correia
Rose Dagul
Kim Kate
Helena Hamilton
DJs
Alien Jams
Team GBH

Graham Dunning – *Rhythm and Drone Set*
http://grahamdunning.com/

Dunning, the master of Mechanical Techno and Ghost in the Machine music, presents his delicate and accidental techno, assembled from discarded turntables and "junk" yet entirely capable of making the dancefloor move. Graham is currently "Maker in Residence" at the Machines Room (http://machinesroom.org/) researching, and hosts a show on NTS (lhttp://www.ntslive.co.uk/about). Read more about his work on 'Mechanical Techno' here in a recent FACT profile: http://www.factmag.com/2015/04/25/graham-dunning-mechanical-techno/

Iris Garrelfs
http://irisgarrelfs.com/

Iris Garrelfs is a composer/performer intrigued by change, fascinated by voices and definitely enamoured by technology. She often uses her voice as raw material which she transmutes into machine noises or choral works. Pieces and performances have been featured in exhibitions, festivals and as part of exhibitions internationally, including her most recent 'performance walk' at the National Gallery as part of the Soundscapes Late series. Aside from her music making, Iris is the co-founder of underground playground and test tube for new sound Sprawl (http://www.sprawl.org.uk/)

Nuno Correia
https://soundcloud.com/coden

Nuno is a researcher and audiovisual artist, interested in interactive multi-sensorial experiences. Nuno's work has been presented in more than 20 countries. Currently, he is a researcher at Goldsmiths, University of London (EAVI Group), working on the project Enabling Audiovisual User Interfaces (AVUIs). Nuno will present AVZones, an iPad audiovisual app – an audio sequencer/looper with a visualizer. AVZones is open source and work in progress, built with openFrameworks and Maximilian, and part of the Enabling AVUIs research project. More info: http://www.nunocorreia.com/portfolio/avzones

When & Where

Amersham Arms
388 New Cross Road
SE14 6TY London
United Kingdom

Thursday, October 8, 2015 at 8:00 PM - Friday, October 9, 2015 at 2:00 AM (BST)

Add to my calendar

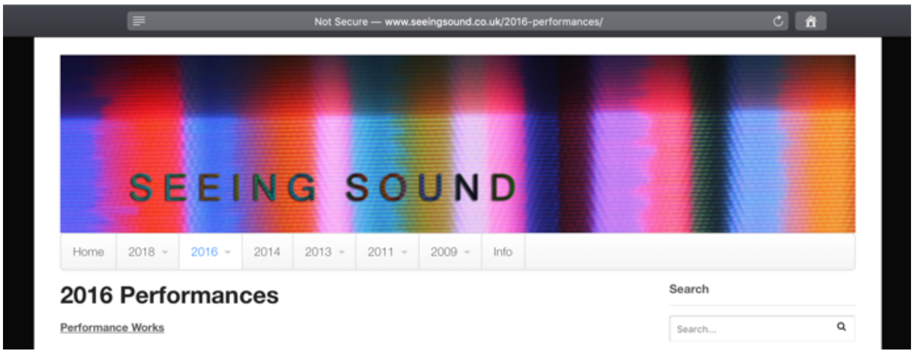
Organizer

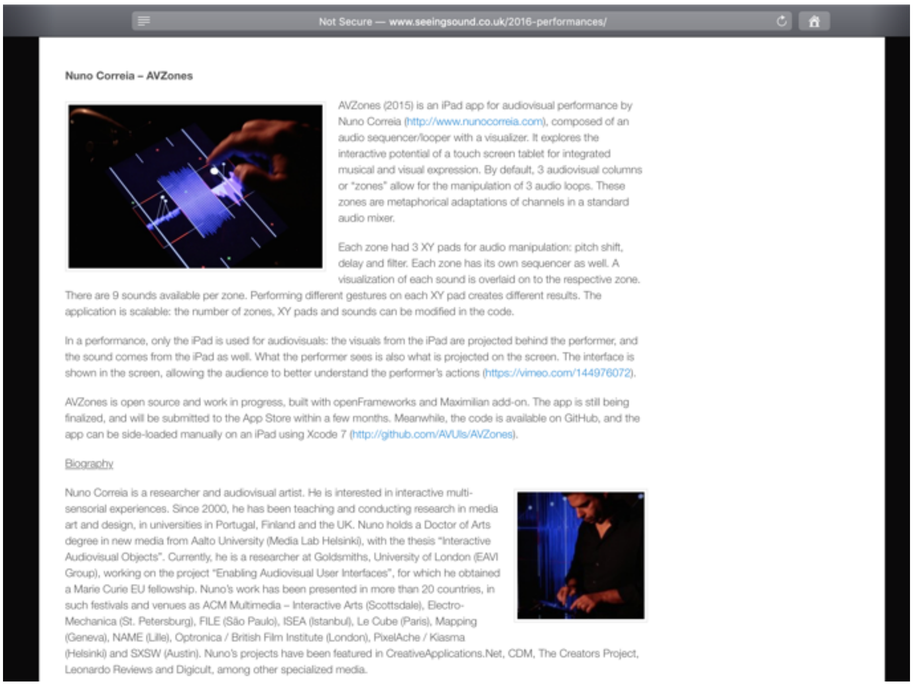
Contact the Organizer
View organizer profile

16 past events on Eventbrite


EAVI XIII
Graham Dunning
Iris Garrelfs
Nuno Correia
Rose Dagul
Kim Kate
Helena Hamilton
DJs ~
Chloe Frieda
(Alien Jams/NTS)
Team GBH
8th October
Amersham Arms
SE14 6TY
£5
8pm till late
eavixiii.eventbrite.com

Photos and promotional materials: solo performance, Seeing Sound Symposium, Bath, 9/4/2016





Not Secure — www.seeingsound.co.uk/2016-performances/

Nuno Correia – AVZones

AVZones (2015) is an iPad app for audiovisual performance by Nuno Correia (http://www.nunocorreia.com), composed of an audio sequencer/looper with a visualizer. It explores the interactive potential of a touch screen tablet for integrated musical and visual expression. By default, 3 audiovisual columns or "zones" allow for the manipulation of 3 audio loops. These zones are metaphorical adaptations of channels in a standard audio mixer.

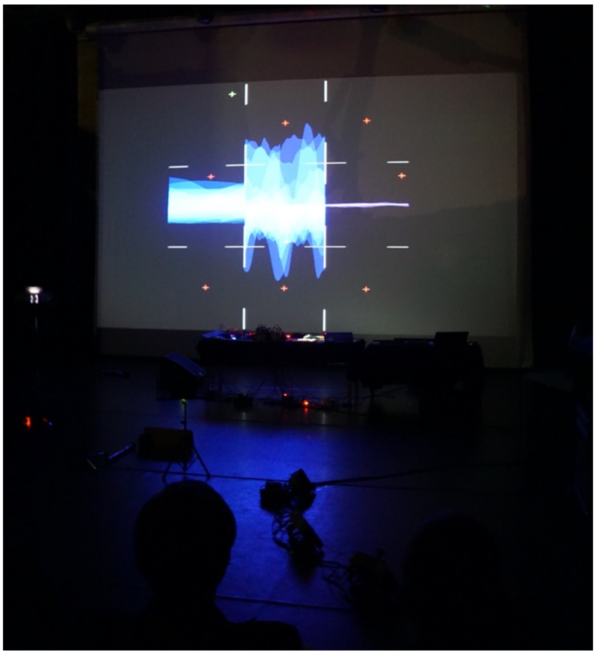
Each zone had 3 XY pads for audio manipulation: pitch shift, delay and filter. Each zone has its own sequencer as well. A visualization of each sound is overlaid on to the respective zone. There are 9 sounds available per zone. Performing different gestures on each XY pad creates different results. The application is scalable: the number of zones, XY pads and sounds can be modified in the code.

In a performance, only the iPad is used for audiovisuals: the visuals from the iPad are projected behind the performer, and the sound comes from the iPad as well. What the performer sees is also what is projected on the screen. The interface is shown in the screen, allowing the audience to better understand the performer's actions (https://vimeo.com/144976072).

AVZones is open source and work in progress, built with openFrameworks and Maximilian add-on. The app is still being finalized, and will be submitted to the App Store within a few months. Meanwhile, the code is available on GitHub, and the app can be side-loaded manually on an iPad using Xcode 7 (http://github.com/AVUIs/AVZones).

Biography

Nuno Correia is a researcher and audiovisual artist. He is interested in interactive multi-sensorial experiences. Since 2000, he has been teaching and conducting research in media art and design, in universities in Portugal, Finland and the UK. Nuno holds a Doctor of Arts degree in new media from Aalto University (Media Lab Helsinki), with the thesis "Interactive Audiovisual Objects". Currently, he is a researcher at Goldsmiths, University of London (EAVI Group), working on the project "Enabling Audiovisual User Interfaces", for which he obtained a Marie Curie EU fellowship. Nuno's work has been presented in more than 20 countries, in such festivals and venues as ACM Multimedia – Interactive Arts (Scottsdale), Electro-Mechanica (St. Petersburg), FILE (São Paulo), ISEA (Istanbul), Le Cube (Paris), Mapping (Geneva), NAME (Lille), Optronica / British Film Institute (London), PixelAche / Kiasma (Helsinki) and SXSW (Austin). Nuno's projects have been featured in CreativeApplications.Net, CDM, The Creators Project, Leonardo Reviews and Digicult, among other specialized media.



Talks

Promotional materials: Talk at Goldsmiths, University of London, 18/6/2014

From: **Atau Tanaka** atau@goldsmithsdigital.com
Subject: Nuno Corriea talk 18 June
Date: 11 June 2014 at 08:22
To: EAVI eavi@goldsmithsdigital.com

Introduction to Audiovisual User Interfaces

18 June 2014, 4pm - 5.30pm
Goldsmiths, Richard Hogarth Building, room 308

Abstract

Nuno Correia will present the research project "Enabling Audiovisual User Interfaces" that he is conducting at EAVI (Embodied AudioVisual Interaction group), Goldsmiths, University of London. It is a 2-year project, supported by a Marie Curie EU fellowship. The main research question is: how can interconnected sound and image be used to create more usable, accessible, playful and engaging user interfaces? To address this issue, a new UI paradigm is proposed – AVUI (AudioVisual User Interface). AVUI links interaction, sound and image, building upon the concept of Graphical User Interface (GUI) by adding interconnected sound and image. The research hypothesis is: the introduction of AVUI, integrating interrelated sonic and visual feedback, reacting to user interactions, will lead to more usable, accessible, playful and engaging UIs, as compared to a traditional GUI – particularly in use cases where accessibility and/or engagement are determinant. During the talk, the project will be contextualised with related research. The methodology and expected outcomes will be presented. Opportunities for collaboration with related projects will also be discussed.

Bio

Nuno Correia is a researcher, media artist and musician. He is interested in enabling interactive multi-sensorial experiences. Since 2000, he has been teaching and conducting research in media art and design, in universities in Portugal, Finland and the UK. Nuno holds a Doctor of Arts degree in new media from Aalto University (Media Lab Helsinki), with the thesis "Interactive Audiovisual Objects", and an MSc in innovation management from the Technical University of Lisbon. Currently, he is a researcher at Goldsmiths, University of London (EAVI group), working on the project "Enabling Audiovisual User Interfaces", for which he obtained a Marie Curie EU fellowship. Nuno's work has been presented in more than 20 countries.

EAVI - Embodied AudioVisual Interaction
Goldsmiths, University of London
http://eavi.goldsmithsdigital.com
http://www.nunocorreia.com

Promotional materials and press: Talk at Sound Bureau Katowice, 31/1/2015

slaskie.naszemiasto.pl/rusza-biuro-dzwieku-katowice-2015-zobacz-co-sie-bedzie/ar/c13-3254616

Rusza Biuro Dźwięku Katowice 2015. Zobacz, co się będzie działo

AG • 23 stycznia 2015 • Zaktualizowano 28 lutego 2017, 11:55

Rusza Biuro Dźwięku Katowice 2015 *materiały prasowe*

31 stycznia w Strefie Centralnej w Katowicach ruszy Biuro Dźwięków Katowice 2015.

Zobacz: Power of Trinity zbiera pieniądze na nowy teledysk. Czy fani pomogą?

Biuro Dźwięku Katowice to projekt, który ma na celu integrację aktywnych i początkujących artystów zainteresowanych muzyką elektroniczną. W ramach projektu odbyły się warsztaty muzyczne, które poprowadzili Piotr Ceglarek, Jan Dybała, Kuba Sojka i Artur Oleś.

- Katowice są miastem muzyki awangardowej. Cieszy się ona dużą popularnością wśród Ślązaków, o czym świadczy m.in. Tauron Nowa Muzyka. Twórczość alternatywna się rozwija, dlatego postanowiliśmy stworzyć warsztaty, które umożliwią amatorom poznanie tajników muzyki elektronicznej - mówi Łukasz Kałębasiak, rzecznik prasowy Instytucji Kultury Katowice-Miasto Ogrodów.

Program:

godz. 18.30
Nuno Correia – wykład-prezentacja pt. "Audiovisual
Nuno Correia - artysta, muzyk i badacz nowych mediów. Podczas prezentacji opowie o swoich doświadczeniach zawodowych na przestrzeni 10 lat. W pierwszej części opowie o działaniach z Video Jackiem, co ma bezpośredni związek z tematem jego pracy doktorskiej oraz książki pt. "Interactive Audiovisual Objects". Druga część poświęcona będzie pracy z narzędziami kreacji audiowizualnej: AudioVisual Vector Exchange i niedawno prowadzonych badaniach nad AudioVisual User Interfaces - AVUIs

godz. 20.00
Łukasz Kacperczyk – koncert
Warszawski twórca elektroakustycznej muzyki eksperymentalnej opartej przede wszystkim na dźwiękach generowanych przez syntezator modularny. Nie stroni jednak od instrumentów wykonanych własnoręcznie. Interesuje go niepowtarzalność i ulotność efektów pracy z takim zestawem oraz jego ogromny potencjał brzmieniowy.

godz. 21.00
Kaeesy Kaeesy – live act

Zobacz też:

Najnowsze

Izrael - Polska - 1:2: zobacz najlepsze MEMY PO MECZU

Śląskie lekarki pozują i przypominają kobietom o badaniach

Galeria wstydu. Te plakaty wyborcze wiszą nielegalnie

NADĘTY BRZUCH? Zgubisz go w 3 dni. Wystarczy rano...
Reklama

Punkty i nagrody za recykling

Elastyczne oferty i pełna swoboda – usługi telekomunikacyjne coraz bardziej nowoczesne
Materiał informacyjny

Konkurs „Fundusz Naturalnej Energii"

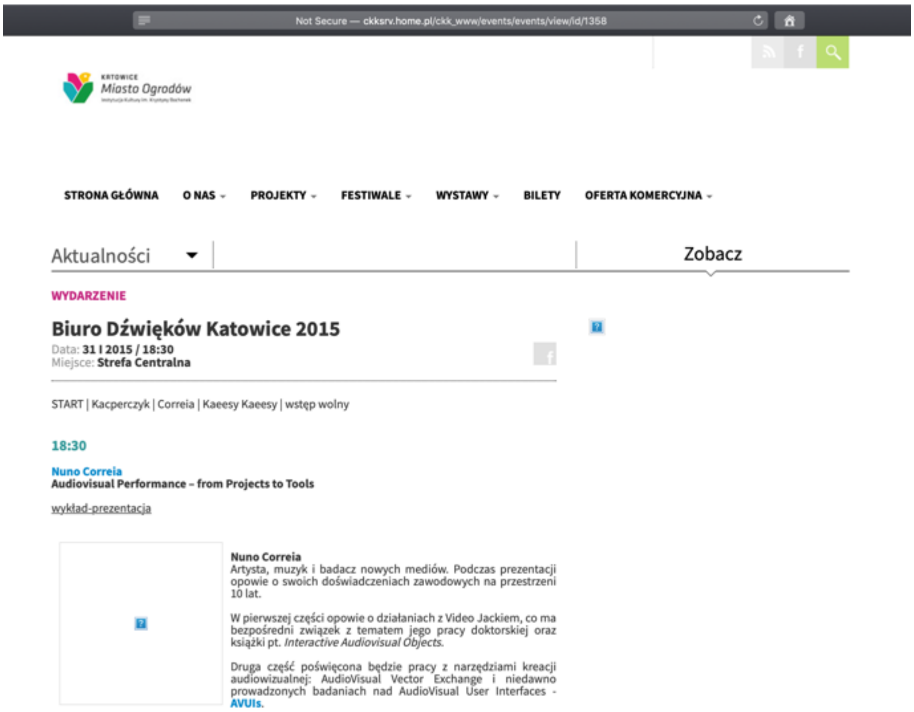
Edukacja, prewencja, dopłaty. Jak walczymy ze smogiem?

Tony śmieci na cmentarzach. Kontenery przepełnione ZDJĘCIA

Od tego zależy, jaka będzie Twoja emerytura
Materiał partnera zewnętrznego PZU

BLACK
Black Friday 2019 w Media Expert. Najciekawsze

Not Secure — falatrzecia.blogspot.com/2015/01/biuro-dzwieku-katowice-2015-start.html
TRZECIA FALA
WYDARZENIA
INTERFERENCJE
TEKSTY
RELACJE
TRZECIA FALA
KONTAKT
Szukaj a znajdziesz...
Home » BIURO DŹWIĘKU KATOWICE » KAEESY KAEESY » ŁUKASZ KACPERCZYK » NUNO CORREIA » STYCZEŃ 2015 » Biuro Dźwięku Katowice 2015 START! ::: KACPERCZYK / CORREIA / KAEESY KAEESY
Biuro Dźwięku Katowice 2015 START! ::: KACPERCZYK / CORREIA / KAEESY KAEESY
KONCERT / WYKŁAD / IMPREZA
31.01.2015 I Strefa Centralna I Katowice
start: 18:30 / wstęp: FREE
NUNO
CORREIA
KACPERCZYK
KAEESY KAEESY
/link FB/
:::: godz. 18.30 – wykład-prezentacja pt. "Audiovisual Performance – from Projects to Tools"
☆ NUNO CORREIA
Artysta, muzyk i badacz nowych mediów. Podczas prezentacji opowie o swoich doświadczeniach zawodowych na przestrzeni 10 lat. W pierwszej części opowie o działaniach z Video Jackiem, co ma bezpośredni związek z tematem jego pracy doktorskiej oraz ksążki pt. "Interactive Audiovisual Objects". Druga część poświęcona będzie pracy z narzędziami kreacji audiowizualnej: AudioVisual Vector Exchange i niedawno prowadzonych badaniach nad AudioVisual User Interfaces - AVUIs.
Master and Margarita - The Seventh Proof - On Stage @ PixelAche
from Video Jack
03:04
vimeo
ARTYŚCI / PROJEKTY
WYDAWNICTWA
INFO / INICJATYWY
ZINY / BLOGI / MAGAZYNY
STACJE I AUDYCJE RADIOWE
MIEJSCA
TRZECIA FALA / FACEBOOK
Trzecia Fala
Like Page
3.8K likes
INTERFERENCJE #8
WIDMA
INTERF...
SOUNDCLOUD
47:12
Cookie policy
::: ALIAŻ ::: MARZEC
WIDMA
::: ALIA...
SOUNDCLOUD
30 TRACKS
Cookie policy
::: LONG ALIAŻ ::: MARZEC



Not Secure — ckksrv.home.pl/ckk_www/events/events/view/id/1358

KATOWICE Miasto Ogrodów

STRONA GŁÓWNA | O NAS | PROJEKTY | FESTIWALE | WYSTAWY | BILETY | OFERTA KOMERCYJNA

Aktualności

Zobacz

WYDARZENIE

Biuro Dźwięków Katowice 2015

Data: **31 I 2015 / 18:30**
Miejsce: **Strefa Centralna**

START | Kacperczyk | Correia | Kaeesy Kaeesy | wstęp wolny

18:30

Nuno Correia
Audiovisual Performance – from Projects to Tools

wykład-prezentacja

Nuno Correia
Artysta, muzyk i badacz nowych mediów. Podczas prezentacji opowie o swoich doświadczeniach zawodowych na przestrzeni 10 lat.

W pierwszej części opowie o działaniach z Video Jackiem, co ma bezpośredni związek z tematem jego pracy doktorskiej oraz książki pt. *Interactive Audiovisual Objects*.

Druga część poświęcona będzie pracy z narzędziami kreacji audiowizualnej: AudioVisual Vector Exchange i niedawno prowadzonych badaniach nad AudioVisual User Interfaces - **AVUIs**.

katowice.wyborcza.pl/katowice/1,35063,17338834,Moj_pomysl_na_weekend__Moze_tez_skorzystasz_.html?disab

wyborcza.pl
KATOWICE
Niedziela 17.11.2019 Dzisiejsze wydanie papierowe

Wiadomości z Katowic

Mój pomysł na weekend. Może też skorzystasz?

MICHALINA BEDNAREK 31 stycznia 2015 | 00:00

Jak według uczestników warsztatów Biuro Dźwięku brzmią Katowice, mogli posłuchać m.in. uczestnicy ubiegłorocznej Industriady (Fot. Radosław Kaźmierczak / Materiały Instytucji Kultury Ars Cameralis)

NAJCZĘŚCIEJ CZYTANE
KATOWICE Górnicy z JSW będą protestować wspólnie z hutnikami z ArcelorMittal. Pikieta w Krakowie
Toyota umacnia sojusz z Subaru. Napęd na cztery koła w hybrydach i elektrykach
WARSZAWA Warszawa ma największą w Unii sieć ciepłowniczą. Tak powstaje ciepło w stolicy
WARSZAWA Sieć ciepłownicza w Warszawie rozrasta się o kolejne kilometry. Co roku podłączane jest do niej ponad 200 budynków

Jeśli jeszcze nie wiesz, jak spędzić weekend może skorzystasz z tych podpowiedzi?

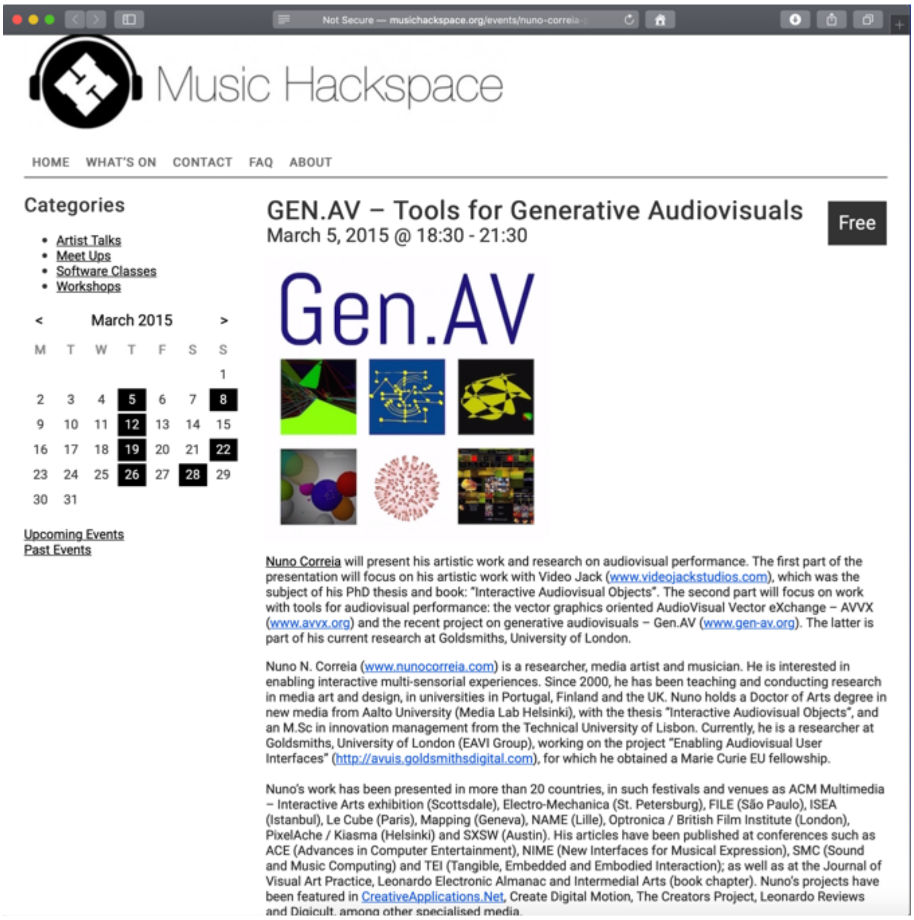
Maciek Bywalec, perkusista w zespole Fair Weather Friends

W sobotę zapewne zawitam na warsztaty katowickiego Biura Dźwięku połączone z koncertem. Na początek o godz. 18.30 w Strefie Centralnej wykład "Audiovisual Performance - from Projects to Tools" wygłosi Nuno Correia, artysta, muzyk i badacz nowych mediów.

Organizatorzy zapowiadają, że podczas prezentacji opowie o swoich doświadczeniach zawodowych z ostatnich 10 lat. Pokaże też, jak pracować z narzędziami kreacji audiowizualnej. Dla mnie, jako muzyka, to szansa na poszerzenie horyzontów. Potem chętnie zostanę na koncercie Łukasza Kacperczyka, warszawskiego twórcy elektroakustycznej muzyki eksperymentalnej opartej przede wszystkim na dźwiękach generowanych przez syntezator. Co jednak dla mnie ciekawsze, ten artysta gra także na instrumentach, które sam wykonał. A na deser pochodzący z Będzina beatmaker oraz raper - Kaeesy Kaeesy. Potem wybieram się do Inqubatora w starym katowickim dworcu. Tam wystąpi The Phantom.


www.facebook.com/events/783827828320761/
Search
Nuno
Home
Create
Events
Events
Calendar
10
Birthdays
Discover
Hosting
+ Create Event
KACPER
CZYK +
CORREIA
+KAEESY
K A EESY
JAN
31
Biuro Dźwięku Katowice 2015 START | Kacperczyk | Correia | Kaeesy Kaeesy |
Public · Hosted by Biuro Dźwięku Katowice
Going
Saturday, January 31, 2015 at 6:30 PM UTC+01
More than a year ago
Strefa Centralna
Plac Sejmu Śląskiego 2, 40-032 Katowice, Poland
Show Map
About
Discussion
106 Went · 39 Interested
Share this event with your friends
Details
godz. 18.30
Nuno Correia – wykład/prezentacja pt. "Audiovisual Performance – from Projects to Tools" (www.nunocorreia.com)
Nuno Correia - artysta, muzyk i badacz nowych mediów. Podczas prezentacji opowie o swoich doświadczeniach zawodowych na przestrzeni 10 lat. W pierwszej części opowie o działaniach z Video Jackiem, co ma bezpośredni związek z tematem jego pracy doktorskiej oraz książki pt. "Interactive Audiovisual Objects". Druga część poświęcona będzie pracy z narzędziami kreacji audiowizualnej: AudioVisual Vector Exchange i niedawno prowadzonych badaniach nad AudioVisual User Interfaces - AVUIs (avuis.goldsmithsdigital.com).
Wykład/prezentacja prowadzony będzie w języku angielskim.
godz. 20.00
Łukasz Kacperczyk – koncert
Warszawski twórca elektroakustycznej muzyki eksperymentalnej opartej przede wszystkim na dźwiękach generowanych przez syntezator
Related Events
See More
Crux 5.4 – audio + video live...
6 PM at Five Miles
244 guests
Interested · Going
Björk | The O2, London
Tuesday at The O2
11,842 guests
Interested · Going
Robert Henke: CBM 8032 AV
Thu Jan 23, 2020 at Barbican ...
134 guests
Interested · Going
Sunday Afternoon Swing at ...
3:30 PM at Jamboree
50 guests
Interested · Going
Modern Panic X
11 AM at Truman Brewery, F Bl...
84 guests
Interested · Going
Alva Noto & Anne-James Ch...
8 PM at LSO St Luke's
by Barbican Centre
Interested · Going
English (US) · Português (Portugal) · Polski · Español · Français (France)
Privacy · Terms · Advertising · Ad Choices · Cookies · More
Facebook © 2019

Promotional materials: Talk at Music Hackspace, London, 5/3/2015



Not Secure — musichackspace.org/events/nuno-correia-

Music Hackspace

HOME WHAT'S ON CONTACT FAQ ABOUT

Categories

- Artist Talks
- Meet Ups
- Software Classes
- Workshops

< March 2015 >

M	T	W	T	F	S	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

Upcoming Events
Past Events

GEN.AV – Tools for Generative Audiovisuals

March 5, 2015 @ 18:30 - 21:30

Free

Gen.AV

Nuno Correia will present his artistic work and research on audiovisual performance. The first part of the presentation will focus on his artistic work with Video Jack (www.videojackstudios.com), which was the subject of his PhD thesis and book: "Interactive Audiovisual Objects". The second part will focus on work with tools for audiovisual performance: the vector graphics oriented AudioVisual Vector eXchange – AVVX (www.avvx.org) and the recent project on generative audiovisuals – Gen.AV (www.gen-av.org). The latter is part of his current research at Goldsmiths, University of London.

Nuno N. Correia (www.nunocorreia.com) is a researcher, media artist and musician. He is interested in enabling interactive multi-sensorial experiences. Since 2000, he has been teaching and conducting research in media art and design, in universities in Portugal, Finland and the UK. Nuno holds a Doctor of Arts degree in new media from Aalto University (Media Lab Helsinki), with the thesis "Interactive Audiovisual Objects", and an M.Sc in innovation management from the Technical University of Lisbon. Currently, he is a researcher at Goldsmiths, University of London (EAVI Group), working on the project "Enabling Audiovisual User Interfaces" (http://avuis.goldsmithsdigital.com), for which he obtained a Marie Curie EU fellowship.

Nuno's work has been presented in more than 20 countries, in such festivals and venues as ACM Multimedia – Interactive Arts exhibition (Scottsdale), Electro-Mechanica (St. Petersburg), FILE (São Paulo), ISEA (Istanbul), Le Cube (Paris), Mapping (Geneva), NAME (Lille), Optronica / British Film Institute (London), PixelAche / Kiasma (Helsinki) and SXSW (Austin). His articles have been published at conferences such as ACE (Advances in Computer Entertainment), NIME (New Interfaces for Musical Expression), SMC (Sound and Music Computing) and TEI (Tangible, Embedded and Embodied Interaction); as well as at the Journal of Visual Art Practice, Leonardo Electronic Almanac and Intermedial Arts (book chapter). Nuno's projects have been featured in CreativeApplications.Net, Create Digital Motion, The Creators Project, Leonardo Reviews and Digicult, among other specialised media.

Photos and promotional materials: Talk at Alexandrinsky Theatre, New Stage, St. Petersburg, 20/3/2015

Promotional materials: Talk at Queen Mary University of London, 18/8/2015

Subject: Next Instruments and Interactions meeting, Tuesday 18th at 3 PM
Date: Mon, 17 Aug 2015 10:41:04 +0100
From: Oussama Metatla <o.metatla@qmul.ac.uk>
To: people@qmedia.qmul.ac.uk <people@qmedia.qmul.ac.uk>, c4dm@eecs.qmul.ac.uk <c4dm@eecs.qmul.ac.uk>, mat-students-12@eecs.qmul.ac.uk, mat-students-13@eecs.qmul.ac.uk, Mat-students-11@eecs.qmul.ac.uk, Mat-students-10@eecs.qmul.ac.uk, eecs-mat-students-14@qmul.ac.uk, cogsci@eecs.qmul.ac.uk
CC: o.metatla@qmul.ac.uk

Hi,

For the next Instruments and Interactions meeting, Nuno Correia from Goldsmiths University will present a talk on "Enabling Audiovisual User Interfaces", see abstract and bio below.

The talk is from 3pm to 4pm in room G2 in the Engineering Building.

All welcome. Please feel free to forward this on.

Best,
Oussama.

Abstract

Nuno Correia will present the research project "Enabling Audiovisual User Interfaces" that he is conducting at EAVI (Embodied AudioVisual Interaction group), Goldsmiths, University of London. It is a 2-year project, started in mid-2014 and supported by a Marie Curie EU fellowship. The main research question is: how can interconnected sound and image be used to create more usable, accessible, playful and engaging user interfaces? To address this issue, a new UI paradigm is proposed – AVUI (AudioVisual User Interface). AVUI links interaction, sound and image, building upon the concept of Graphical User Interface (GUI) by adding interconnected sound and image. The research hypothesis is: the introduction of AVUI, integrating interrelated sonic and visual feedback, reacting to user interactions, will lead to more usable, accessible, playful and engaging UIs, as compared to a traditional GUI – particularly in use cases where accessibility and/or engagement are determinant. He will present the main research threads he has been developing as part of his project, partly in collaboration with Queen Mary, University of London. Project link: http://avuis.goldsmithsdigital.com

Bio

Nuno Correia is a researcher, media artist and musician. He is interested in enabling interactive multi-sensorial experiences. Since 2000, he has been teaching and conducting research in media art and design, in universities in Portugal, Finland and the UK. Nuno holds a Doctor of Arts degree in new media from Aalto University (Media Lab Helsinki), with the thesis "Interactive Audiovisual Objects", and an MSc in innovation management from the Technical University of Lisbon. Currently, he is a researcher at Goldsmiths, University of London (EAVI group), working on the project "Enabling Audiovisual User Interfaces". Nuno's work has been presented in more than 20 countries. Homepage: http://www.nunocorreia.com

--
Dr. Oussama Metatla

Postdoctoral Research Fellow
School of Electronic Engineering & Computer Science
Queen Mary University of London
Mile End Road, London E1 4NS, UK

Room: CS321 - Tel: +44 (0) 207 882 7249
Web: http://www.eecs.qmul.ac.uk/~oussama/

Photos and promotional materials: Talk at VJ London , 2/12/2015

Promotional materials and photo: Talk at MiXD Symposium, Birmingham, 16/3/2016

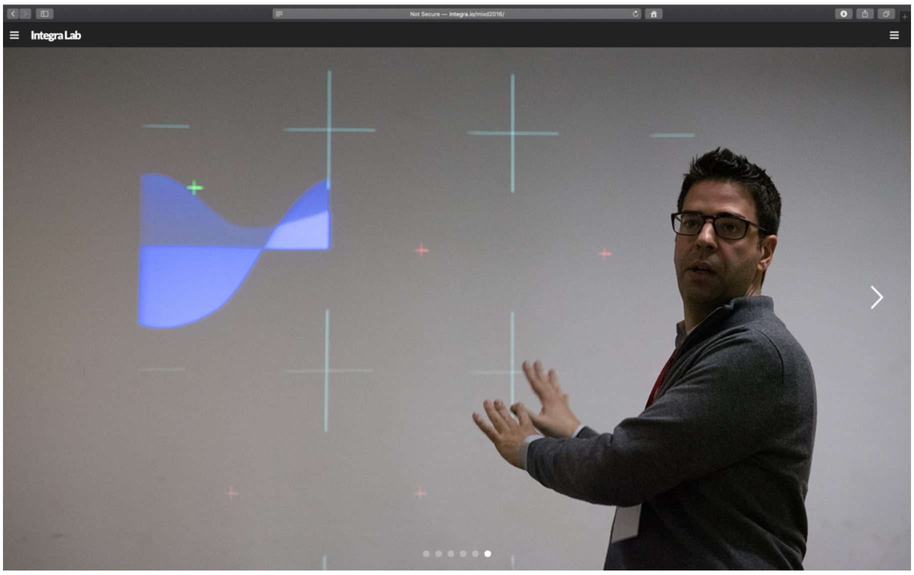
3:30pm – 3:45pm **D AV Zones: a multitouch audio sequencer and visualizer**

Speakers: Nuno Correia New Lecture Hall (Birmingham City University, Paradise Place, Birmingham B3 3HG)

AVZones (2015) is an iPad app for audiovisual performance by Nuno Correia (www.nunocorreia.com), composed of an audio sequencer/looper with a visualizer. It explores the interactive potential of a touch screen tablet for integrated musical and visual expression. By default, 3 audiovisual columns or "zones" allow for the manipulation of 3 audio loops. These zones are metaphorical adaptations of channels in a standard audio mixer. Each zone had 3 XY pads for audio manipulation: pitch shift, delay and filter. Each zone has its own sequencer as well. A visualization of each sound is overlaid on to the respective zone. There are 9 sounds available per zone. Performing different gestures on each XY pad creates different results. The application is scalable: the number of zones, XY pads and sounds can be modified in the code.

In a performance, only the iPad is used for audiovisuals: the visuals from the iPad are projected behind the performer, and the sound comes from the iPad as well. What the performer sees is also what is projected on the screen. The interface is shown in the screen, allowing the audience to better understand the performer's actions (vimeo.com/144976072).

AVZones is open source and work in progress, built with openFrameworks and Maximilian add-on. The app is still being finalized, and will be submitted to the App Store within a few months. Meanwhile, the code is available on GitHub, and the app can be side-loaded manually on an iPad using Xcode 7 (github.com/AVUIs/AVZones).



Promotional materials: Talk at Varvara & Mar Studio, Tallinn, 16/4/2016

GUEST SPEAKER: NUNO CORREIA
at Varvara & Mar Studio

16th April 2016 at 6pm / Pärnu mnt 154, room 312 (3rd floor)
45 minutes presentation + 15 minutes questions

SUMARY TALK

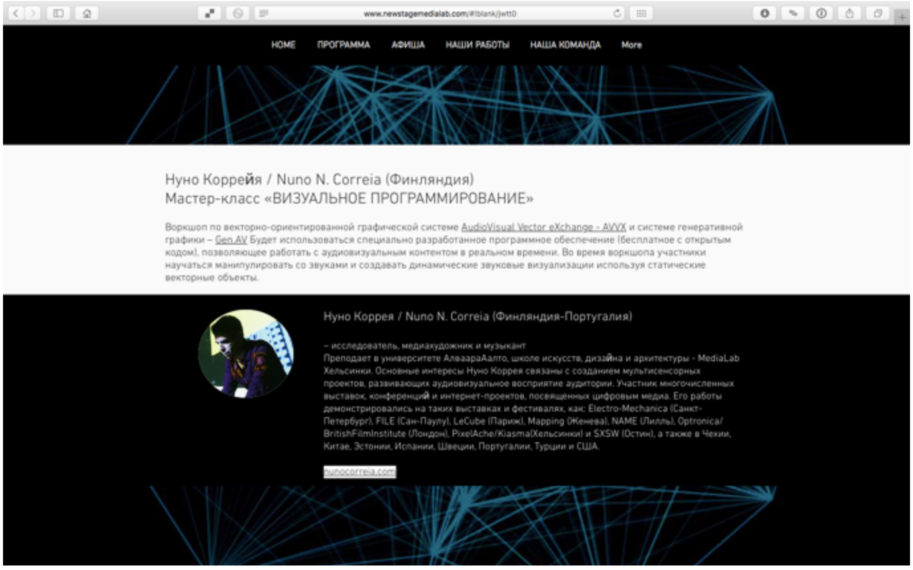
Nuno will present his artistic work and his research on audio-visuals and interaction design. Regarding his own performance work, he will highlight the recent project AVZones (www.avzones.org). He will also present tools for audiovisual performance and interaction design with sound and graphics. Namely, he will discuss his work with a community of audiovisual artists and developers (www.gen-av.org), and a toolkit for interaction design with audiovisuals (ofxAVUI). These projects are part of his current research at Goldsmiths, University of London.

ABOUT
Nuno Correia

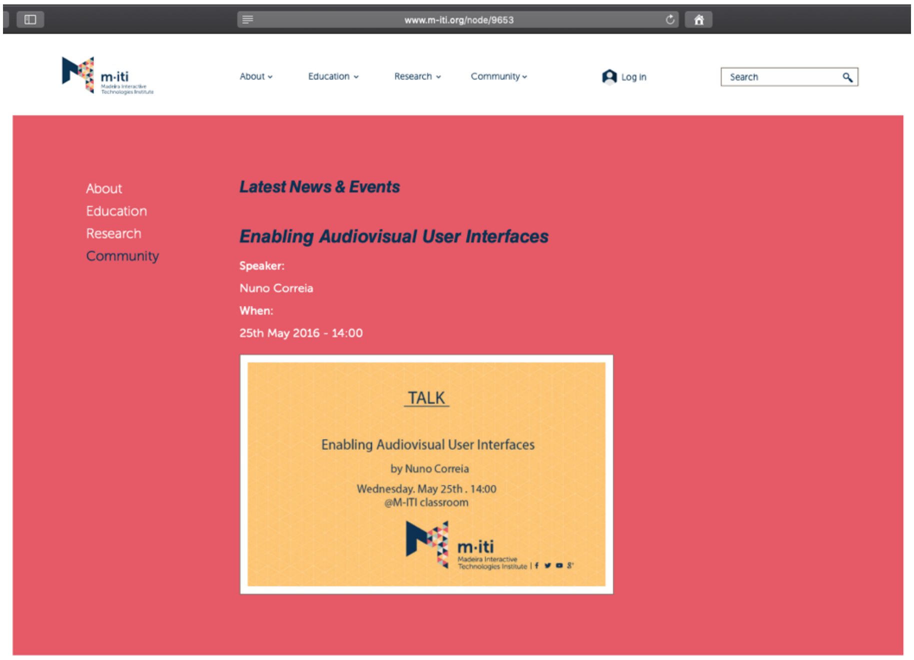
Nuno N. Correia is a researcher, media artist and musician. He is interested in enabling interactive multi-sensorial experiences. Since 2000, he has been teaching and conducting research in media art and design, in universities in Portugal, Finland, Estonia and the UK. Nuno holds a Doctor of Arts degree in new media from Aalto University (Media Lab Helsinki), with the thesis "Interactive Audiovisual Objects", and an M.Sc in innovation management from the Technical University of Lisbon. Currently, he is a researcher at Goldsmiths, University of London (EAVI Group), working on the project "Enabling Audiovisual User Interfaces" (http://avuis.goldsmithsdigital.com), for which he obtained a Marie Curie EU fellowship.

www.nunocorreia.com

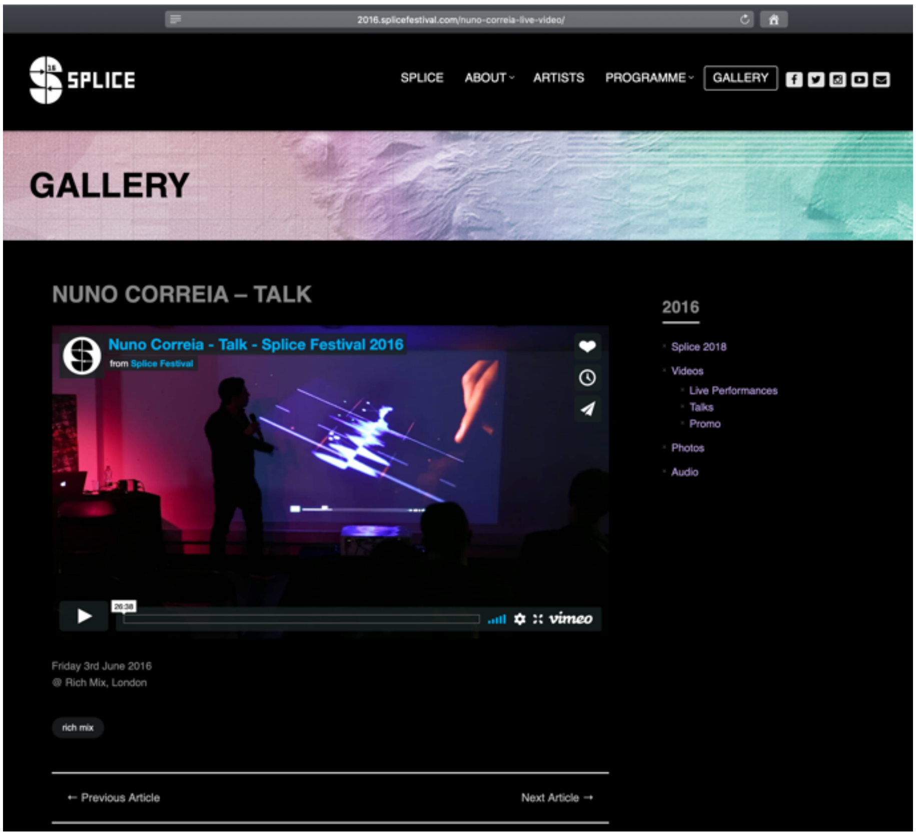
Promotional materials: Master Class at Alexandrinsky Theatre, New Stage, St. Petersburg, 24/4/2016

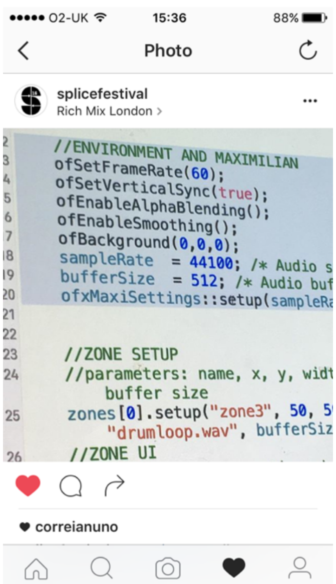


Promotional materials: Talk at Madeira Interactive Technologies Institute, 25/5/2016



Photos and promotional materials: Talk at Splice Festival, London, 3/6/2016

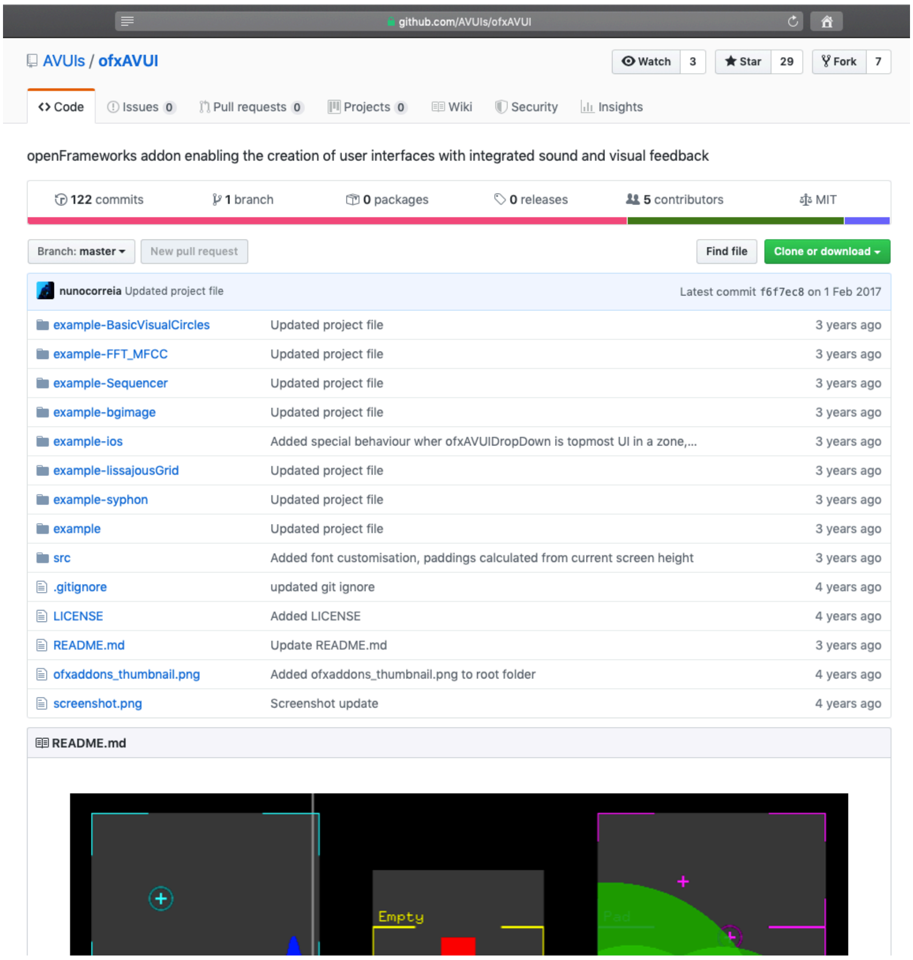




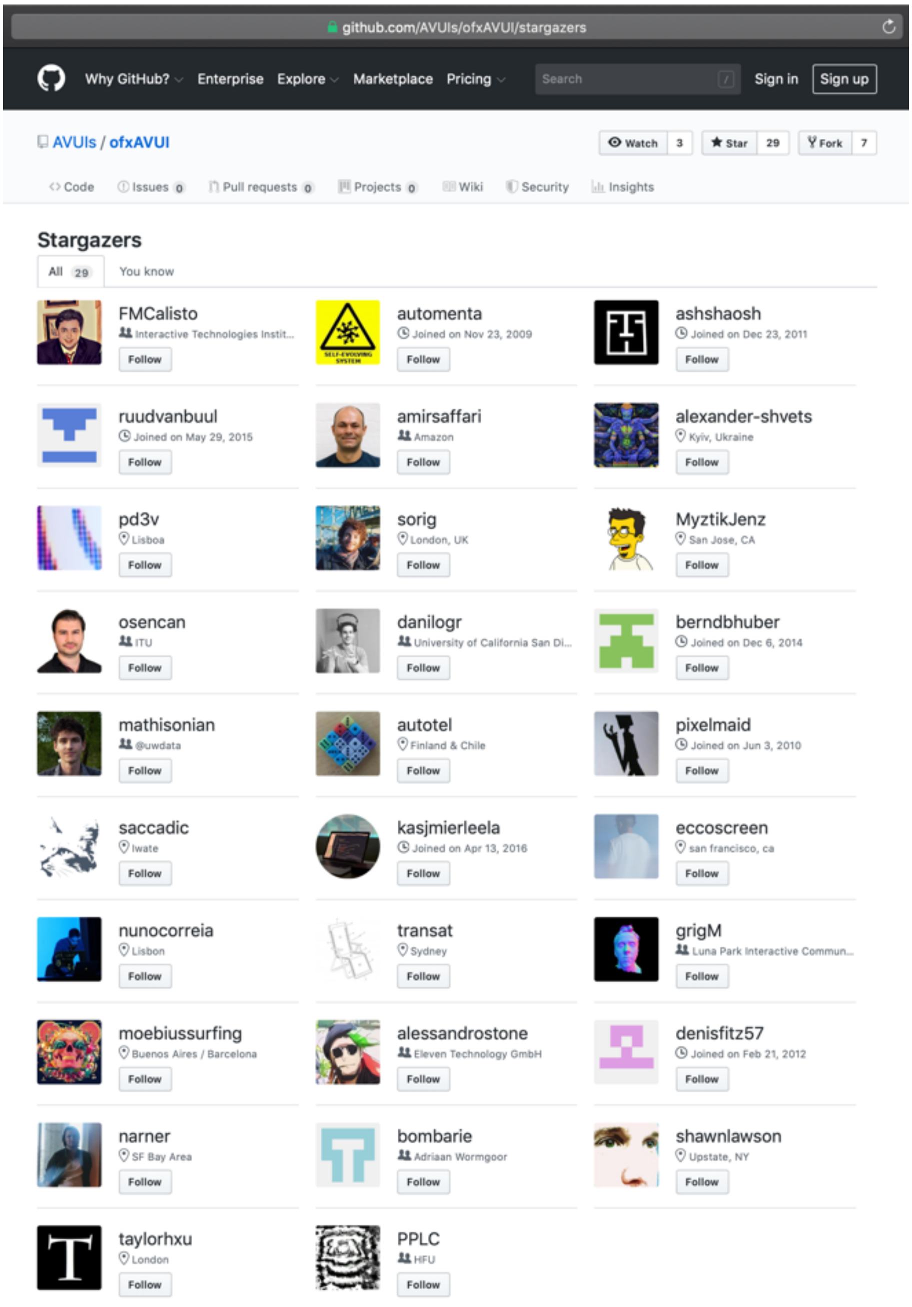
Software

https://github.com/AVUIs

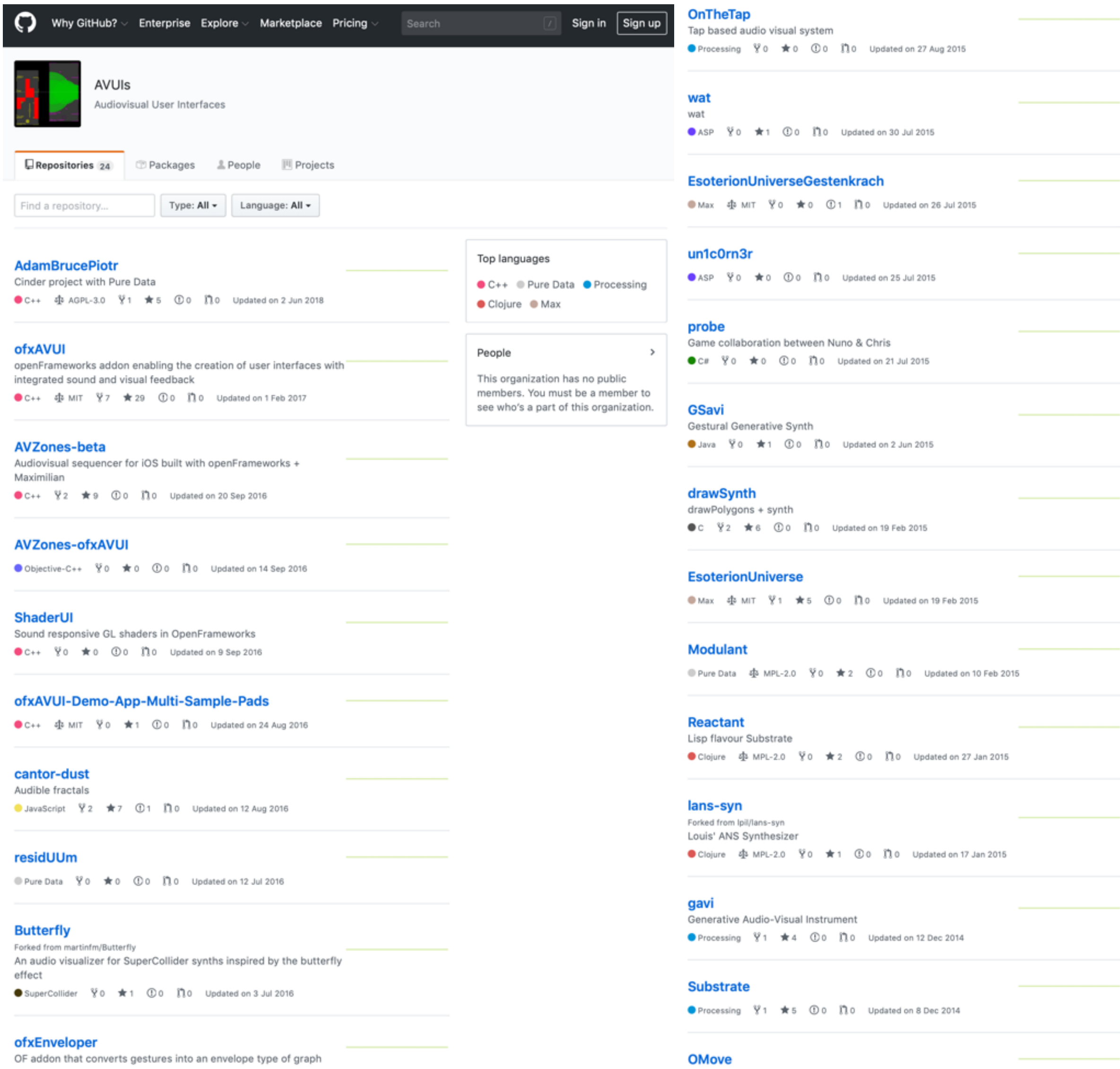
Main repository: ofxAVUI



Followers on the main repository: ofxAVUI



24 repositories in total from hackathon participants



Peer-reviewed Articles

Full papers

Correia, Nuno N. and Tanaka, Atau (2017) *AVUI: Designing a toolkit for audiovisual interfaces.* In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17 . ACM, New York, NY, USA, pp. 1093-1104. ISBN 978-1-4503-4655-9 (doi:https://doi.org/10.1145/3025453.3026042)
http://gala.gre.ac.uk/id/eprint/20966

Correia, Nuno N., Castro, Deborah and Tanaka, Atau (2017) *The role of live visuals in audience understanding of electronic music Performances.* In: Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences. ACM Press, New York, NY, USA, pp. 1093-1104. ISBN 978-1-4503-5373-1 (doi:https://doi.org/10.1145/3123514.3123555)
http://gala.gre.ac.uk/id/eprint/20972

Metatla, Oussama, Correia, Nuno N., Martin, Fiore, Bryan-Kinns, Nick and Stockman, Tony (2016) *Tap the shapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface.* In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16 . ACM, New York, NY, USA, pp. 1055-1066. ISBN 978-1-4503-3362-7 (doi:https://doi.org/10.1145/2858036.2858456)
http://gala.gre.ac.uk/id/eprint/20968

Correia, Nuno N. and Tanaka, Atau (2014) *User-centered design of a tool for interactive computer-generated audiovisuals.*In: Proceedings ICLI 2014 2nd International Conference on Live Interfaces. ICLI, Lisbon, pp. 86-99. ISBN 978-989-746-060-9
http://gala.gre.ac.uk/id/eprint/20971

Short papers

Correia, Nuno N. (2016) *AV Zones – Tablet app for audiovisual performance.* In: Proceedings of the 2016 International Conference on Live Interfaces. Experimental Music Technologies (EMuTe) Lab/REFRAME Books, University of Sussex. ISBN 978-0-9931996-8-4
http://gala.gre.ac.uk/id/eprint/20969

Correia, Nuno N. and Tanaka, Atau (2015) *Prototyping audiovisual performance tools: A hackathon approach.* In: NIME Proceedings of the International Conference on New Interfaces for Musical Expression. The School of Music and the Center for Computation and Technology (CCT), Louisiana State University, Baton Rouge, Louisiana. ISBN 978-0-692-49547-6
http://gala.gre.ac.uk/id/eprint/20965

AVUI: Designing a Toolkit for Audiovisual Interfaces

Nuno N. Correia
Madeira-ITI, U. Madeira
Funchal, Portugal
nuno.correia@staff.uma.pt

Atau Tanaka
Goldsmiths, University of London
London, UK
a.tanaka@gold.ac.uk

ABSTRACT
The combined use of sound and image has a rich history, from audiovisual artworks to research exploring the potential of data visualization and sonification. However, we lack standard tools or guidelines for audiovisual (AV) interaction design, particularly for live performance. We propose the AVUI (AudioVisual User Interface), where sound and image are used together in a cohesive way in the interface; and an enabling technology, the ofxAVUI toolkit. AVUI guidelines and ofxAVUI were developed in a three-stage process, together with AV producers: 1) participatory design activities; 2) prototype development; 3) encapsulation of prototype as a plug-in, evaluation, and roll out. Best practices identified include: reconfigurable interfaces and mappings; object-oriented packaging of AV and UI; diverse sound visualization; flexible media manipulation and management. The toolkit and a mobile app developed using it have been released as open-source. Guidelines and toolkit demonstrate the potential of AVUI and offer designers a convenient framework for AV interaction design.

ACM Classification Keywords
H.5.2 User Interfaces: Auditory (non-speech) feedback; H.5.2 User Interfaces: Graphical user interfaces (GUI); H.5.2 User Interfaces: Prototyping; H.5.2 User Interfaces: User-centered design

Author Keywords
User interface; toolkit; interface builder; audiovisual; crossmodal interaction; prototyping; participatory design; interaction design; hackathons.

INTRODUCTION
The combination of audio with image has a long tradition, from color organs used by the composer Scriabin in the early 20th century [27] to the pioneering computer graphics explorations of John Whitney in the 1960s [26]. The advent of powerful personal computers for media manipulation, from the 1990s, gave further impulse to audiovisual (AV) performance [32].

These artistic explorations mirror how the brain deals with multi-sensorial information. Research on sensory substitution has explored how the brain replaces functions of one sense by another [1]. The perception of flashing lights can be manipulated by sound: a single flash of light can be seen as consisting of two flashes if displayed simultaneously with multiple sound signals [34]. In the McGurk effect, the perception of an auditory phoneme changes depending on the image [24]. An important factor for crossmodality is congruency – non-arbitrary associations between different modalities. A congruent AV display can result in better performance and higher engagement than arbitrary associations between sound and image [25].

Sound and image have been studied in HCI in different application areas, including accessibility in assistive displays [11], improvement of task accuracy in driving [31], enjoyability and performance in games [5, 25]. Despite the potential for facilitating usability and engagement, there is a lack of design guidelines and standard tools for AV interaction design. Specifically, current solutions for AV performance that facilitate UI integration with content are laborious to implement, and lack aesthetic concerns regarding coherence of UI and visuals. Interface design for AV performance is mostly subjective, and there are no established best practices. These best practices would benefit performers, audience, software developers, interaction designers, researchers and students.

We propose the AVUI (AudioVisual User Interface) where the interaction of sound and image in the interface extends the concept of GUIs. We seek to 1) leverage practices in AV performance for sketches and prototypes, using participatory design methods; 2) implement best practices into a consolidated prototype; 3) propose guidelines and a software toolkit to allow designers to easily integrate sound and image in the UI of future systems and products.

This paper reports on the multi-stage design, development, release and evaluation of a software toolkit, ofxAVUI. We present related work; the participatory design and qualitative methods used; and their results; followed by AVUI guidelines proposed, discussion and conclusions.

RELATED WORK
AV Performance and Tools
AV performance combines live manipulation of sound and image [4]. This distinguishes it from VJing (Video Jockey performance), where a visual performer accompanies a musician [8]. A number of artists are concerned with creating interfaces and systems for AV performance. Levin developed

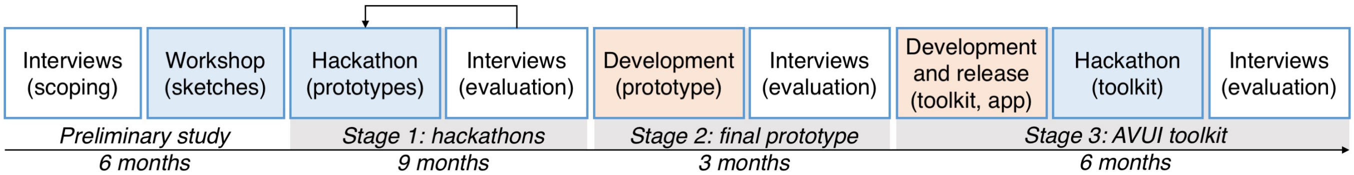


Figure 1. Summary of the four cycles of the research: preliminary study; stages 1, 2 and 3. Workshops and hackathons in blue, interviews in white and development phases in orange. Stage 1 had two iterations

painterly interfaces for audiovisual performance [20]. Magnusson uses abstract GUIs to represent musical structures [23]. Iwai creates playful pieces, crossing genres between game, installation and performance [29].

Most commercial VJ software, such as Modul8 (http://www.modul8.ch), focus on video playback and manipulation, with limited generative graphics capabilities, and only "fairly low-level musical features" [33]. AV performers "often rely on building their own systems" [33] with coding frameworks such as openFrameworks (OF) (http://openframeworks.cc). Therefore, an important element of VJing and AV performance is the use of Do It Yourself (DIY) tools [32]. This requires "a high level of technical ability on the part of the user" [33].

Solutions such as Processing (http://processing.org) offer sound toolkits, which contain visualization modules (ex: Minim), and also offer GUI modules (ex: Control P5). Similar examples could be given for other creative coding environments. But although these UI and sound/visualization modules can be combined, this is laborious, and they are not integrated out of the box, neither functionally nor aesthetically. They do not offer easy ways to implement an "AV+UI" solution, nor a GUI designed to be integrated with visuals. We aim to provide an ease of implementation, and high level of integration.

AV Systems and Interaction Design

Schofield et al. created Cinejack for "directing narrative video through live musical performance" [33], in collaboration with artists. In MelodicBrush, a user-centered design approach is adopted to design a tabletop AV system linking calligraphy and music [15]. Wiethoff and Gehring created an interactive media façade system through an iterative approach: key data collection; user research; data analysis; design concepts; and experience concepts [35]. These studies have used multi-stage, user-centered approaches, which we adopted in our work. However, they each only evaluated one system, making generalization difficult. We apply these methods to evaluate a large number of projects in order to glean design insights across multiple systems.

The New Interfaces for Musical Expression (NIME) community has been active in studying the combination of visuals, sound and interaction design. The work of Jordà, such as FMOL and Reactable, is relevant due to the interplay of interaction and sound visualization strategies [17]. The authors of residUUm, one of the prototypes resulting from the current research, have presented their performance approaches for AV generation and audience visibility [30]. Rouages is another AV project that is concerned with audience understanding [2].

Hook has developed an interface for VJ performances, Waves [14]. He also studied VJing from the viewpoint of the performer [13] and identified three main themes: *aspirational*, *live* and *interaction*. Within *interaction*, he identifies the following sub-themes: *constraining interactions*; *haptically direct*; *parallel interaction*; *immediacy*; *manipulable media*; *reconfigurable interfaces*; and *visible interaction*. While he focuses on video content, he recognizes the need for generative media tools. We will use Hook's framework for our study, strengthening the potential generalizability of the work by building upon prior qualitative findings.

METHODS

The research took place over a 2-year period and involved a preliminary study and three main stages (Figure 1): Pre) scoping interviews and brainstorming workshop; 1) hackathons for prototype development; 2) a "consolidation prototype" gathering best practices and expert evaluation; 3) creation of a software development toolkit for facilitating integration of AV with UI. The toolkit was released as open-source, tested in a hackathon and with internet users, and further tested by rebuilding the final prototype from Stage 2 as a publicly released product using the toolkit. The iterative cycle enabled "a dynamic process of invention, distributed across events" [9]. Hackathons and interviews (and subsequent thematic analysis) were used as methods throughout the studies.

Hackathons

Hackathons are coding events in DIY communities where "small teams produce working software prototypes in a short time period" and these events are often centered around a common theme or technology [18]. The hackathon challenge is an important part of the method. It sets a common task in a motivating way to participants, making a hackathon "a moment of design" [16]. In a hackathon, solutions are "conceived in response to those challenges" [21]. These elements make it a fun, easy to understand technique for participatory design and code development.

Thematic Analysis

We conducted thematic analyses of interviews in the three studies, based on techniques in [3]. We coded the responses based on emerging patterns and issues arising, then collated the codes into potential themes. We used Hook's themes as a basis for our coding. We complemented this theoretic analysis approach with an inductive analysis independent of any pre-existing coding frame. This allowed us to build on prior work, contribute our new insights, and achieve a balanced, thorough and inclusive structure of main themes and sub-themes.

PRELIMINARY STUDY: INTERVIEWS, BRAINSTORMING

In our preliminary study [7], we conducted interviews with 12 audiovisual performers, asking them about their practice, their tools, and their needs and desires as performers. The analysis of the interviews brought forth a series of key issues: modularity, flexibility and reconfigurability; ease of hardware/software integration; instrument-like expressivity and fluidity; integration of environmental elements; generative capabilities and diversity; communication of process to the audience; reliability and speed. These concepts on the whole match and confirm the issues identified by Hook under the theme *interaction*. Generative capabilities and diversity connects to their forward looking theme of *aspiration*, and the need for a visual equivalent to sound synthesizers. The 12 interviewees provided us with a group of experts that we would consult throughout the different studies of the research – the evaluators of Stages 1 and 2 were from this same group.

The ideas from the interviews then informed a brainstorming workshop, with 19 participants (including two from the previous interview stage). The one-day workshop structure was comprised of two parts: the first one adopting the "bootlegging" idea generation technique [12]. For part 2, we extended this with a more focused, structured re-examination of ideas from part 1, which we called "Re-boot".

The five breakout groups produced five sketches (storyboards and wireframes) of procedural audiovisual performance tools. Two sketches, *Gestural Touchscreen* and *Meta/Vis*, were particularly successful in addressing the challenges set out in the workshop. Both rely on the expressive potential of multitouch interaction, employing different solutions for reconfigurability: the former allows for loading and manipulating vector graphics, and the latter adopts a simplified data-flow mechanism. Project descriptions and sketches are seen at `http://www.gen-av.org/sketches/`.

STAGE 1: PARTICIPATORY DESIGN OF PROTOTYPES

Hackathon and Hack Challenge

Using the key themes and the sketches from the preliminary study as input and inspiration, we ran two hackathons in an iterative cycle, Gen.AV 1 and Gen.AV 2. The objective was to leverage knowledge from AV performers into prototypes combining AV and UI, where best practices could be adopted in a future AVUI toolkit. We sent out a call for participation, with coding knowledge as prerequisite, and interviewed applicants. Each hackathon took place over two days.

Both Gen.AV 1 and 2 followed the same structure: 1) Introduction: a presentation on the previous stages of the study and results achieved so far, goals and structure of the workshop; 2) Conceptualization and sketching; and 3) Software development. 23 participants took part in Gen.AV 1 (five female and 18 male). Gen.AV 2 had 13 participants (two female and 11 male), three of whom had taken part in Gen.AV 1. We divided participants into five (Gen.AV 1) and six (Gen.AV 2) groups, distributing prior programming experience evenly across groups.

We created hack challenges based on key conclusions (*in parenthesis*) from the preliminary study [7]. They were: 1) computer-generated sound and visuals (*generative capabilities and diversity*); 2) powerful and fluid manipulation – "like an instrument" (*instrument-like expressivity and fluidity*); 3) single-screen – what the performer sees is what the audience sees (*communication of process to the audience*); and 4) possibility to reconfigure the interface (*modularity, flexibility and reconfigurability*). The resulting projects were presented in two public performances. Five projects were showcased in the Gen.AV 1 performance, and six in Gen.AV 2. Each group produced a 10 minute performance.

Projects

We present the five projects from Gen.AV 1. *ABP* is an animation engine and sound visualizer, where the user can define color, geometry and animation parameters. In *drawSynth*, a GUI allows controlling sound and image – users can draw vector shapes and select colors, which are sonified by a synthesis engine. *Esoterion Universe* consists of a 3D space that can be filled with planet-like audiovisual objects, each containing a GUI to modify their visual and sonic properties. *GS.avi* is an instrument that generates continuous spatial visualizations and music from the gestural input of a performer. *Modulant* allows for drawing images, using paintbrush type of tools, which are then sonified.

Six projects were built during Gen.AV 2. *Butterfly* is an audio visualizer which allows for the combination and control of four audio synthesizers, by means of manipulating icons distributed in four XY pads on the screen. *Cantor Dust* generates, displays, and sonifies Cantor set type fractals as sound and visuals. *EUG* further develops *Esoterion Universe* from Gen.AV 1, adding 3D gestural control with a Leap motion sensor. *OnTheTap* plays with the tactile, analog feel of tapping surfaces as interaction input, captured as audio. *residUUm* allows for the creation and manipulation of AV particles, with a variable lifespan, by clicking and dragging on the screen. *Wat* creates a chaotic 3D texture based on cellular automata (Figure 2). All the projects were uploaded to GitHub for download or source-code modification, accessible from `http://www.gen-av.org`. In addition to the code, the project descriptions are available from the same link, facilitating running and replicating the projects.

Expert Interviews

After the performances, the projects were tested for ease of installation and robustness. Six projects were chosen: *Esoterion Universe*, *GS.avi* and *Modulant* from Gen.AV 1; and *Butterfly*, *residUUm* and *Wat* from Gen.AV 2. These were evaluated by expert reviewers, established audiovisual artists who had taken part in the preliminary study interviews. Each expert was given two projects for review, and at least one week time to practice with the software. Thus, each project was evaluated twice (project evaluators E1 and E2). We then conducted semi-structured interviews with the reviewers, lasting an average of 15 minutes per project. They served to follow up on the key issues emerging from the preliminary stage. The questions addressed: 1) the AV content and the relationship between sound and image; 2) ease of use of the software; 3) fluidity, AV manipulability and behavior as an "instrument"; 4) flexibility and reconfigurability of the interface; 5) potential usefulness for other artists and performers.

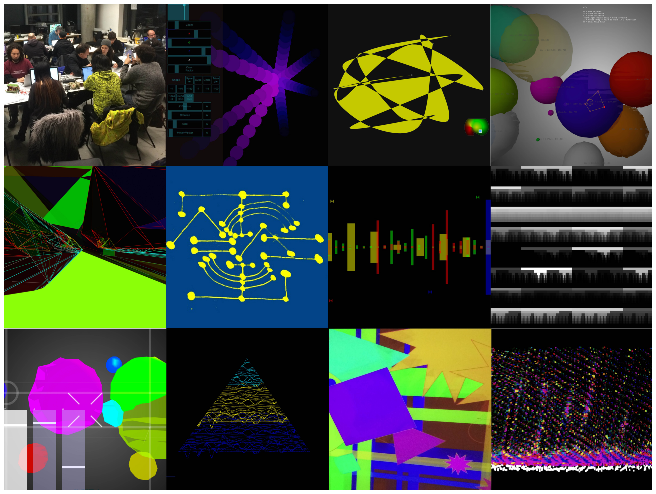


Figure 2. Stage 1: Projects from the hackathons – left to right, then top to bottom: hackathon, ABP, drawSynth, Esoterion Universe, GS.avi, Modulant, Butterfly, Cantor Dust, EUG, OnTheTap, residUUm, Wat.

We conducted a thematic analysis of the interviews. From this analysis, three main themes emerged: *Experience*, *Interfaces* and *Media*. We retained these themes for our analyses of the different stages of the research. Our starting point were Hook's themes related to *interaction*: *constraining interactions* (importance of constraints and focus); *haptically direct* (physical connection); *parallel interaction* (simultaneous control of multiple parameters); *immediacy* (immediate response from the software); *manipulable media* (powerful and varied manipulation of media); *reconfigurable interfaces* (reorganize controls to fit a particular performance); and *visible interaction* (make interaction visible to an audience) [13]. *Immediacy* was a pre-requisite for the selection of the projects. *Constraining interaction* was not detected. We decided to merge *haptically direct* and *parallel interaction*, as they would appear combined in our data. We split *interaction* into two new main themes: *experience* and *interface*. Additional themes related to content emerged, originating the new main theme *media*.

Experience

Visible interaction: Two of the evaluators, in their own practice, prefer not to convey the interface to the audience, and wish to have a separate screen with the GUI for the performer (GS.avi, E1; Modulant, E2), for three main reasons: 1) the visual output could be re-routed without GUI to other software for additional manipulation (GS.avi, E2); 2) the interface is something the audience may not want to see; and 3) having a separate screen would allow for a more complex interface for the performer (Wat, E2). Another evaluator is interested in showing the UI to the audience and "conveying the performer's control on-screen" as part of the experience (residUUm, E2). Taking that approach would allow for "visually reflecting that agency onto the screen", making it understandable for an audience (Butterfly, E2).

Haptic and parallel: The reviewers confirmed the desire for interaction that "provides a sensation akin to being in direct contact or touching and molding media" [13]. The projects did not allow for either haptic or parallel interaction. Multitouch tablets and hardware controllers were mentioned as means to achieve a parallel interaction. Compatibility with hardware controllers and tablets is desired to achieve parallel interaction: controllers would allow for the physicality and "flexibility of an instrument" (Esoterion Universe, E1) and interaction "in a tracking pad on the computer is confusing", with a tablet-based approach being suggested (GS.avi, E2).

Object-oriented: When auditory and visual domains are combined, audiovisual objects can emerge, if simultaneity and a plausible common cause occur [19]. Three projects followed an object-oriented approach, by grouping audio and visual content into distinguishable entities (*Esoterion Universe*, *Butterfly* and *residUUm*). In the first two cases, a GUI was overlaid on the visuals for continuous manipulation. In the third project the opportunity for manipulation occurs only at the genesis of the object. The object approach was considered as being fruitful (Esoterion Universe, E2).

Interface

Reconfigurable interfaces: The reconfiguration of UI becomes possible only by editing the code, which requires specific technical knowledge (Esoterion Universe, E1). Some projects organized the code in order to make it easier to reconfigure: "it's very easy to add your own synths" (Butterfly, E1).

Interface mappings: In some instances, a complex one-to-many mapping of interface to media parameters was considered successful (Butterfly, E2). Evaluators felt that more parameters should be controlled from the UI, resulting in insufficient mapping (Butterfly, E2; Wat, E2). Scalability of layout is desired, as it would allow for additional UI elements (Butterfly, E1).

Interface clarity: The lack of a parameter space in UI elements – an indication of the parameter range, and the current status – was considered problematic in some projects (residUUm, E2).

Interface aesthetics: The gestural aspect of one project was seen as innovative and appealing (GS.avi, E1). It was suggested that it could become more integrated in the visuals by visualizing the gestures (GS.avi, E1). The visual design of the interface is considered to be even more important when conveyed to an audience. In some cases, this design was considered to be unappealing to be shown (Modulant, E2). One reviewer considers that the UI itself should be dynamic, animated in response to the sound (Butterfly, E2). One of the projects adopts a logic of interactive quadrants with XY pads, which was considered to be original and clear (Butterfly, E1).

Media

Manipulable media: Some projects were considered to produce outcomes with a narrow range of diversity (Esoterion Universe, E2; Butterfly, E1). In several cases, the projects rely on 2D or 3D spatial metaphors. There is a desire for an expandable canvas or scene where the media can be presented in and navigated through. This is considered to be missing on one project (Modulant, E2) and praised for its implementation on another (Wat, E1, E2). The satisfactory manipulation possibilities of some projects lead them to be considered "instruments" because of their fluidity (Wat, E1; Modulant, E1).

Generative media: Different evaluators appreciated different degrees of randomness. The generative aspect of some projects was considered to be too chaotic (GS.avi, E1; residUUm, E2). In the balance between generative elements and control, the latter is seen as the priority. But a certain degree of randomness is desirable, and considered to be missing in some cases (Modulant, E1).

Media management: The option to load files in some of the projects is appreciated (GS.avi, E2; Modulant, E1). Runtime loading of content is desired (GS.avi, E1). Real-time sharing of media between applications in the same device, using utilities such as Syphon (http://syphon.v002.info) is wished for (Esoterion Universe, E1). The possibility of accessing networked content is also suggested (Esoterion Universe, E2).

Audience Study

In order to study audience understanding of the performers' actions, we asked audience members of the two performances to fill in a questionnaire about the different projects from Gen.AV 1 and 2 (with the exception of *DrawSynth*, a last minute addition). Respectively 45 and 34 respondents answered the questionnaire. The question asked was: "Did you find the connection between the performer's actions and the audiovisual result understandable?", on a scale of 1 to 5. Projects *Esoterion Universe* and *Modulant* from Gen.AV 1, and *Butterfly*, *EUG* and *residUUm* from Gen.AV 2 obtained the best results (*Modulant* with a median of 5, the others with a median of 4). The five projects that achieved the best results make visible both the interface and the parameter space. *Cantor Dust* and *OnTheTap*, both with a median of 3, implement only one of these aspects (visibility of parameter space in *Cantor Dust*) or only temporarily show them (*OnTheTap*). The remaining projects, with a median of 2, employ neither. These observations informed our design principles.

STAGE 2: AV ZONES CONSOLIDATION PROTOTYPE

Prototype Design

The results from Stage 1 fed into design guidelines for a final prototype, an iPad app for AV performance entitled *AV Zones*. It has been released as open-source (https://github.com/AVUIs/AVZones-beta). It adopts the object-oriented concept of "zones": rectangular areas that incorporate UI elements producing and manipulating sound, and a visualization of that sound. The app has three vertical zones, each with three XY pads for audio manipulation, controlling: pitch shift, delay and filter. Each zone has a sequencer, which can record touch information and visualize it. There are nine sounds available per zone, which can be switched at runtime, and replaced in the code. Different touch inputs create different results: tapping for triggering sounds; touch movement for manipulating the sound; two-finger tap to switch on and off; and double tap to trigger special function – menu or sequencer (Figure 3). The app was developed using the OF environment and the Maximilian audio library (https://github.com/micknoise/Maximilian). Both are open-source and cross-platform.

Initial Tests - Performances

We tested *AV Zones* in "real world" settings: four public performances and two demos in conferences. In a performance, only an iPad is used for audiovisuals. What the performer sees is also what is projected to the audience. The interface is shown on the screen, with touch points being represented by white circles. We made minor improvements between performances. For example, a sequencer was added due to the difficulty of interacting simultaneously with the three zones, and the need to automate some of the processes by recording them.

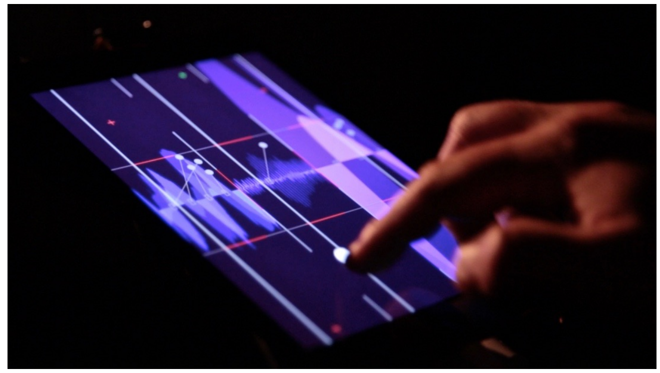


Figure 3. Stage 2: AV Zones prototype.

Expert Interviews

We followed a similar evaluation procedure than in Stage 1: we installed *AV Zones* in iPads and handed them to three evaluators (E1, E2, E3) from our initial expert group. The interviewees tried the app for at least one week. We then ran semi-structured interviews lasting on average 30 minutes. We used the same questions as Stage 1. We ran a thematic analysis, maintaining the three main themes: *media*, *interface* and *experience*. More sub-themes emerged: *sound visualization* (within *media*); *constraining interaction* – a theme that had existed in Hook's analysis but had not appeared in Stage 1; and *playfulness* (within *experience* main theme).

Experience

Constraining interaction: Two evaluators were satisfied with the prototype's design constraints and minimalism, stating that "it's nice to have limitations", having a "minimal simplicity" was pleasing, and its "reduced nature" made it "very appropriate for a live tool" (E1, E3).

Visible interaction: One evaluator would like to be able to hide the UI, completely or partially, and added that by separating what the audience and the performer see, more UI elements could be added on the performer side (E1). Another respondent is satisfied that the audience can see what the performer is doing, as touch points are highlighted with white circles, and would like to see more interactions visualized, such as sound effect manipulation (E2).

Haptic and parallel: One respondent was satisfied with the number of zones and simultaneous control elements (E2), whereas others would like to add an external hardware MIDI controller (E1) or another tablet, creating a dual setup (E3).

Object-oriented: One respondent was pleased with the notion of zones and the way they operate, stating that it was a good concept and design (E2).

Playfulness: The application was considered to be playful – one respondent mentioned multiple times that he enjoyed playing with it, that it was "fun" (E1).

Interface

Interface aesthetics: One evaluator was very pleased with the interface aesthetics of the sequencer functionality, stating that it looks like a "visual music composition" (E1). Another respondent considers that more work could be done in terms of visualizing additional processes in the software, such as loading or choosing sounds (E2).

Interface clarity: The prototype suffers from hidden discoverability issues. Several of the functionalities are activated by different types of touch interaction, not apparent in the UI. Testers had problems activating these, despite the documentation provided (E1, E2, E3). Evaluators complained that it was hard to understand what to do next (E2) leading to getting occasionally stuck (E1).

Reconfigurable interfaces: One respondent in particular was interested in adding reconfigurability options, such as allowing for extending functionalities with software "plugins" that others could build. Another suggestion was having the possibility of grouping zones and nesting them - this would facilitate scaling of zones without overcrowding the screen (E1). One of the respondents suggested adding a back end with substantial configuration options (E3).

Media

Manipulable media: One respondent considers that the software is "a really useful live tool" (E3). The two audio effects, delay and filter, were considered well chosen, and having three simultaneous sounds allows for "enough scope" to maintain a performance (E3). The sequencer is considered an important element for this, as it allows to automate one zone while interacting with others (E1, E2, E3). The prototype is considered to allow for "a different way of approaching sound", less "musical" and "kind of weird" (E1). To have a broader and more musical appeal, two evaluators consider that a stricter timing or "clock" would be important (E1, E3). Having "more authorship and a sense of control" (E1) over the sound is desired. On the visual side, respondents would also like to have greater control. Only one visualization type, with "very little visual configurability" (E3) is considered insufficient. Evaluators would like to be able to have other visualizations and be able to make more choices about them (E1, E2).

Media management: All evaluators would like to be able to load sound files. Although this is possible, it requires modifying the code and re-installing the app, which is inconvenient. The possibility to record sounds is also desired (E2). One evaluator would like to be able to route the visuals to other software for further processing (E1).

Sound visualization: One evaluator considers that the approach followed, to visualize the amplitude levels of the audio buffer, was "fascinating" and "very responsive", particularly at slower speeds (E1). The other two consider this approach to be simplistic, as it does not help to "understand anything about the sound" (E2). They would rather have a "perceptually motivated approach" that would bring it closer to the state of the art (E2) and a two-way interaction between sound and image: not just sound visualization, but also visual sonification (E3).

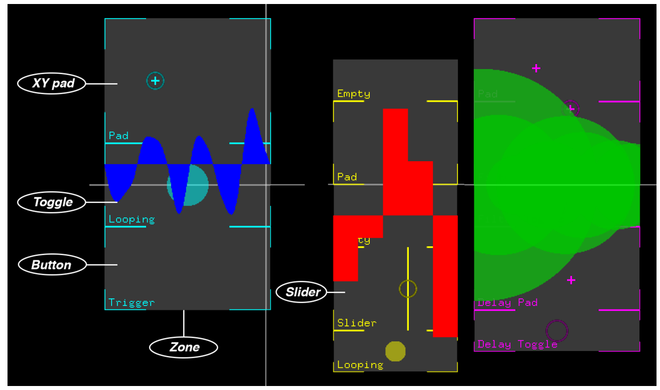


Figure 4. Stage 3: Example of AVUI built with the toolkit, using three zones, with different UI elements and visualizations (explanatory labels in white).

STAGE 3: THE OFXAVUI TOOLKIT

Tookit Design and Development

To assist in making the development of AV work more streamlined, to better integrate interfaces in AV performances, and to make interaction more understandable for audiences, we have developed a toolkit for combining UI with AV content. We generalized knowledge gained from the previous stages in the development of ofxAVUI, a modular, reusable software toolkit to facilitate the production of audiovisual user interfaces. The evaluation of the 11 Stage 1 prototypes and the app from Stage 2 were distilled into a set of design insights, which in turn contributed to the definition of the feature set of our toolkit. This led to a technical specification and software architecture. The design specifications for the toolkit were, divided by main themes: 1) experience – allow for parallel and visible interaction; integrate sound, image and UI following an object-oriented approach; 2) interface – enable reconfigurable interfaces, with flexible mappings; ensure both clarity and aesthetic appeal of interface, harmonized with visuals; 3) media – allow for powerful media manipulation, with procedural content; and adopt a flexible media management. Full design specifications, and their connection with previous stages of the research can be found at `http://www.gen-av.org/avui-design-tables/`.

For the development of the toolkit, we again used OF and Maximilian. We organized the code into three groups of class files: audio, visuals and UI. Each of the three groups has a base class, making it easy to extend and to create a new audio process, a new visualization and a new UI type. The style of the UI is centralized in one class, facilitating the customization of its appearance. It was released as an "add-on" (plug-in for OF), allowing to be integrated in other OF projects by developers. We released the add-on in versions for personal computer and mobile multitouch devices. ofxAVUI was released as open source in our GitHub repository (`https://github.com/AVUIs/ofxAVUI`). As is customary with OF add-ons, we included examples, extensively commented the source code, and adopted the "ofx" prefix. It is now part of the main directory for OF add-ons, in the UI category: http://ofxaddons.com/categories/1-gui.

We kept the object-oriented notion of zones from the app in Stage 2, as an organization structure for combining AV and UI. Each zone has only one sound and one visualization, to reinforce its individuality and its objecthood. Different UI elements can be added to a zone: buttons, toggles, XY pads, sliders, range sliders, drop-down menus and labels. The number of zones can be defined, as well as their: size; position; color palette; UI elements. Any parameter from the UI can be rerouted to any audio feature of the zone, or any other aspect of the software (for example, any graphic on the screen). We kept the minimal UI aesthetics of the prototype. Visualization is an important link between sound and image, therefore we added two visualizations, with more configuration options. We also facilitated the creation of new visualizations, making the visualization module extensible. We incorporated the Syphon protocol, so that media could be channeled, with or without UI, to other applications (Figure 4). These design elements, core to the definition of an AVUI, are exposed to the OF developer through high level function calls, making integration into an OF project straightforward.

As an example of ease of ofxAVUI implementation, only three lines of code are needed to create and configure an AV zone with a button that triggers a sound and associated visualization. UI and visualization inherit the aesthetic properties configured for the zone.

Evaluation

For a first, internal validation of ofxAVUI, we built a general release version of our Stage 2 prototype using the add-on. This new version allowed us to address areas to improve in *AV Zones* identified during Stage 2: interface clarity, media manipulation, and media management. In terms of interface clarity, we separated the multiple functions of the XY pad into dedicated toggles and buttons (on/off toggle, sequencer toggle, sound file drop-down menu). Regarding media manipulation, the sequencer can now record and visualize additional interactions. As for media management, users can add and manage sound files via the Apple iTunes interface, a standard for iOS apps. This final version of *AV Zones* has been released as open source on GitHub (https://github.com/AVUIs/AVZones-ofxAVUI), and can be loaded into an iOS device using Apple's Xcode software. The ofxAVUI add-on allowed us to easily and quickly re-develop our prototype and solve issues detected in Stage 2 (Figure 5).

To evaluate the add-on with other developers, we organized a one-day hackathon to look at its ease of use and effectiveness of development. A call was circulated using the same channels as the Stage 1 hackathons. Eight participants took part in the hackathon (five male, three female). Their profile was similar to the previous participants: audiovisual performers and developers. Four of the participants managed to complete a small project during the one-day event. The projects were: *FFT/MFCC*, audio frequency analyzers and visualizers; *Step Sequencer* for creating rhythmic patterns; *Background Image*, for customizing zones; and *Lisajous and Grid*, two additional

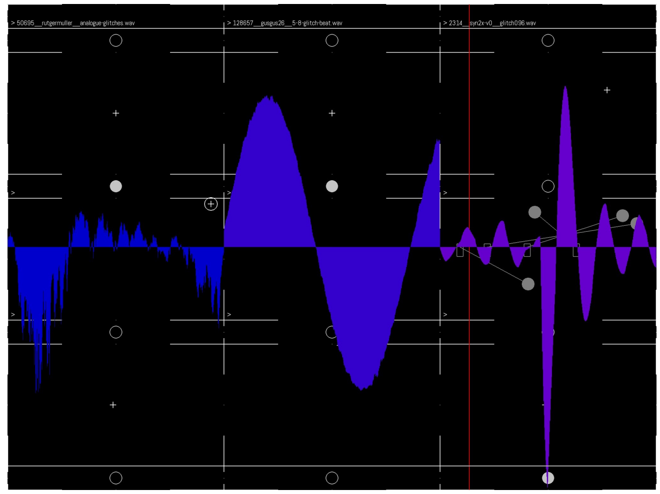


Figure 5. Stage 3: AV Zones, rebuilt with ofxAVUI.

visualizers. These projects expand the toolkit and were added to the ofxAVUI online repository (Figure 6).

In order to obtain further feedback, we reached out to ofxAVUI users on GitHub. Although software downloads are anonymous, 12 individuals had "starred" the repository – a form of following the repository and its updates on GitHub. Of those 12, eight had contact information in their GitHub profiles and were contacted by us. We sent an email asking if they would like to participate in a study. We obtained four replies, and two developers agreed to participate. They developed two projects: a four-zone *Multisampler* and *ShaderUI*, an implementation of sound-responsive shaders. They were also added to our repository.

We conducted face-to-face interviews with the participants in the hackathon (E1-8), and Skype interviews with the two online developers (E9, E10). Interviews lasted on average 30 minutes. The semi-structured interviews addressed: 1) ease of development with ofxAVUI; 2) its usefulness; 3) the appeal of its design; 4) results achieved and satisfaction with those; 5) potential for future use of the add-on.

In our thematic analysis of the interviews, a new main theme emerged: *development*, related to observations on programming and code. From this, three sub-themes were considered: *organization and architecture of code*, *speed and ease of development*, and *patching and building*. This main theme preempted the *reconfigurable interfaces* theme of the previous studies. One additional theme emerged: *scenarios*, under the *experience* main theme.

Development

Organization and architecture of code: Most respondents considered that the code was well organized, with "everything nicely in their respective categories" (E1), and that it was easy to see "how the objects related to each other" (E9). Some evaluators mentioned that the code was easy to extend, as every category has a base class (E3, E10). One respondent highlighted the flexibility in mapping UI parameters to other

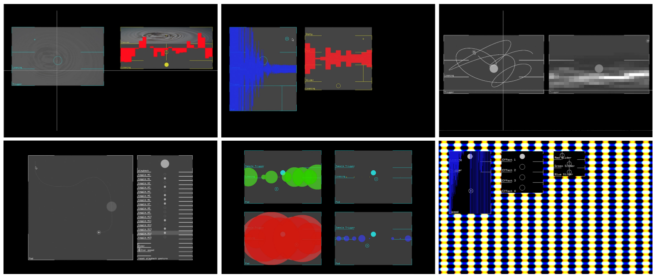


Figure 6. Stage 3: Projects from the ofxAVUI hackathon (first four) and online evaluators (last two) – left to right, then top to bottom: *Background Image*, *FFT/MFCC*, *Lisajous and Grid*, *Step Sequencer*, *Multisampler* and *ShaderUI*.

zones. Different from other UI toolkits, ofxAVUI is, according to our users, designed "to reuse bits and pieces in particular ways to invent new stuff" (E10). Three respondents wanted more abstraction in the code (E2, E8, E9): flexibility regarding multiple types of input (for example, touchscreen or sensors), and to be able to "switch audio engines at will" (E9). One evaluator could not finish the project on time because the toolkit did not support PureData (https://puredata.info) as audio engine (E8). Two evaluators felt that the UI could have taken advantage from the existing ofxGUI toolkit (E1, E3).

Speed and ease of development: ofxAVUI was considered easy to work with by respondents (E3, E4, E7, E9, E10), not just because of its organization and architecture, but also because "it already has the minimum package of sound, UI and visual" built in (E3). One of the respondents considers ofxAVUI easier to use than the two main UI toolkit for OF, ofxUI and ofxGUI (E10). This evaluator considers it easier and faster to prototype with than using related tools, and that "it fills a gap", providing "interesting opportunities that would not be as easily possible previously". Two respondents thought that the add-on could be better documented (E1, E10).

Patching and building: Three respondents would like to have a top-level environment that facilitates the creation of UIs, with a simplified coding language (E1, E3, E7), or by drag-and-drop, as in visual interface builders (E1). Two evaluators expressed interest in having a "master controller" (E5, E10) that could switch zones on and off, and reroute information between them. Another was interested in integrating zones and patching them (E3). Yet another suggested having multiple visualizations per zone, stacked in layers, visualizing different parameters (E5). Two respondents are interested in the implementation of the OSC protocol (http://opensoundcontrol.org) to control other applications and devices (E6, E10), with one stating that it was faster to build a OSC controller with ofxAVUI than with Lemur, a popular tablet controller builder app (E10).

Experience

Visible interaction: In a use case that involves showing ofxAVUI to an audience, such as a performance or a demo, most respondents (E1, E3, E5, E6, E8, E10) consider that revealing the UI is important, making the software "engaging" and "easy to understand". It creates "a more cohesive experience", by showing "the beauty of the internals of the system" (E3). One respondent considers that showing an UI would depend on the use case, and that revealing it "challenges how you interact" in order to "find ways to make audience aware of what's happening" (E5).

Scenarios: Two evaluators considered ofxAVUI well suited to teach sound and visualization (E2, E8). One respondent considered ofxAVUI adequate for game development, particularly pedagogical games (E2). He also suggested that ofxAVUI could be used in more generic applications, for highlighting important tasks. Another evaluator stated that the adoption of OF makes the toolkit more suited to artistic applications, but if made more "portable, or not relying on OF" it could be used for more "day-to-day software development" (E9).

Interface

Interface clarity: Two evaluators requested more visual feedback for changes of state in the UI (E2, E6), such as hovering.

Interface aesthetics: Several respondents liked the minimalist "bare bones" aesthetics of the UI elements (E1, E3, E9). One evaluator wanted to customize UI elements and implement UI "themes", and developed a project for adding background images to zones (E1). One respondent mentioned that the large size of the UI elements "seems more applicable to a touch interface than a mouse interface" (E10).

Media

Sound visualization: One respondent wanted to have more possibilities for audio analysis, and created a project in that direction, based on frequency visualizers. He suggested that more audio information retrieval techniques and 3D visualizations could be added (E2). One respondent wanted to have not only sound visualization, but also visual sonification (E1). Another considered that the visualizations should be used not just for sound, but to visualize other data (E10).

Media manipulation: One evaluator (E4) showed interest in having live audio input. Another respondent (E5) wanted to synthesize sound from an image and its color information. One respondent was interested in rhythmical and quantized aspects of sound, and developed a step sequencer (E3).

AVUI GUIDELINES

The best practices identified in our research allow us to propose the following design guidelines for AVUIs, for use by designers who wish to implement AVUIs, either using ofxAVUI or a different approach. They may be useful for designers who wish to use sound and image together in the interface, either by using ofxAVUI or by using different technologies. These guidelines are divided into three topics, which match the three main themes across the different stages:

1) Maximizing AV Experience

a) Develop AVUIs that can be implemented across multiple platforms and interaction modalities – multitouch interaction seems particularly suited, as it allows for the synaesthetic illusion of touching and molding the audiovisuals;
b) Consider the potential of AVUIs for facilitating visualization of interaction when sharing/showing a screen;
c) Adopt an object-oriented approach, for a harmonious,

coherent and interrelated convergence of audio, image and UI;
d) Facilitate different types of display, allowing for different performer-audience display configurations and hardware.

2) Optimizing Interface Functionality and Aesthetics

a) Use reconfigurable interfaces, possibly with a back-end, that allow to re-map elements of the UI to different sonic features and visual properties; that can also change how the sound is visualized; and that can have an extensible architecture in order to better allow for customization;
b) Explore not simply one-to-one but also one-to-many mappings between UI, audio and visual features;
c) Adopt a minimalist interface aesthetics that integrates well with the visuals, namely regarding color, shape and movement, and that does not detract from the visuals;
d) Reinforce interface clarity by ensuring visibility of all UI elements, their state, parameter space and current position to it; and verify that the visualizations do not detract from this;
e) Allow for hierarchical interfaces, with the possibility of a master control, and communication between modules.

3) Media Strategies

a) Allow for powerful manipulation of sound and image: different forms of media generation, such as different forms of sound and visual synthesis and sampling; multiple audio and visual effects; and experiment with mappings between UI, audio and visuals across different properties;
b) Make use of generative media, due to its variety, flexibility and economy of resources;
c) Try different visualization and sonification approaches, using information retrieval techniques from audio and image;
d) Visualization should reflect not only audio but also the multiple interactions afforded by the UI;
e) Leverage powerful media management features, such as networked content (for example, streamed audio or visuals), audio and visual input, and content sharing between applications in the same device.

DISCUSSION

Our multi-stage research produced a range of concrete outcomes: prototypes by us as well as by participants; a software plug-in toolkit; and an app built with that plug-in. This process allowed us, in an iterative user-centric manner, to gain insight on AVUI design, summarized as a set of AVUI design guidelines above, with implications for design discussed here.

Comparison with Existing Solutions

Since ofxAVUI is built in C++ and with the popular openFrameworks toolkit, it can be easily adopted by digital artists and designers, and embedded in other C++ code (without having to resort to OSC, although it also supports it). Stage 3 evaluators were very positive regarding speed and ease of development with ofxAVUI, compared to other solutions (such as ofxUI and ofxGUI), and most of them consider it easy to use. One of the main trade-offs of ofxAVUI versus other solutions is, due to its inherent pre-packaging of AV and UI, it is not as flexible as using assorted graphics and sound toolkits to build a solution from scratch. On the other hand, it is quicker, and already establishes a harmonization of UI with content.

Object-oriented Integration of Interface with AV

In an AVUI as we propose, sound, visualization and user interface are integrated, functionally and aesthetically, into the same entity. This relates to the concept of audiovisual objects in cognitive science [19]. The results from our studies confirm the appeal of an object-oriented approach to AV interaction design [6]. By analyzing sound and representing it visually in real-time, sound and image are harmonized, synchronous and coherent. Audiovisual congruency is ensured, making use of the identified benefits of crossmodal congruency regarding task accuracy and engagement [5, 11, 25, 31]. This object-oriented approach to AVUI design is apt to situations where the display and interaction plane are fused, as in multitouch displays, allowing for “a sensation akin to being in direct contact or touching and molding media” [13].

Visibility of Interaction

The visibility of interaction is inherent in an AVUI: an interaction triggers either a visual or a sound, which is visualized. An AVUI can be particularly suited for use cases where there are benefits from representing user interaction, such as: performances (our case-study); remote collaboration and telepresence; presentations and demos. However, we have detected different profiles of users regarding visible interaction: some prefer to visually reflect agency onto the screen, “making it understandable for an audience” (Butterfly, E2), others consider that audiences do not necessarily want to see agency on the screen. In functional applications, this will be determined by the task at hand. For creative applications, this can be a matter of taste, as we noted with our expert evaluators in Stage 1. In either case, the AVUI allows the developer to merge visual content with interface elements.

Reconfigurable Interfaces and Flexible Mappings

One of the identified strong points of ofxAVUI was its modularity and reconfigurability: it is designed to recombine UI and media in different configurations and mappings, providing “interesting opportunities that would not be as easily possible previously” (user E10). In ofxAVUI, being able to easily route any UI parameter to any aspect within a zone (sound, visuals, other UI elements) or outside the zone (for example, any graphic on the screen) was considered by our users as innovative with respect to other UI toolkits. Flexibility of mappings between UI, sound and image, and ways to manage that flexibility, are fundamental qualities of an AVUI, and important features in enabling technologies for them to be useful to interaction designers across a range of application domains.

Design Constraints and Speed of Development

User E3 appreciated that ofxAVUI provides the core necessary functionality in sound, UI, and visuals. This is considered an advantage for some: one evaluator considers that he can “prototype a certain part of my process faster” with ofxAVUI, and that it is easier to use than other UI toolkits (E10). For others, ofxAVUI is too constrained precisely because it is tied to specific packages of sound, UI and visuals. These evaluators would like more abstraction, to be able to replace certain elements of the toolkit (for example, the audio library).

Some evaluators would like ofxAVUI to be simpler to use, for example by adding a GUI layer that would allow users to build AVUIs by dragging and dropping elements, as in traditional interface builders. Therefore, there seems to be a desire for both higher level ease of use, and lower level flexibility. A better balance could be pursued in the future between ease of use by pre-packaging elements, and allowing for more architectural flexibility. Having been built with OF is a constraint in itself – OF is popular in media art and design, but not used as much for more generic development. The toolkit could be ported to other frameworks to facilitate its adoption for more generic use cases.

Participatory Design and Hackathons

Our participatory design approach enabled us to leverage artistic knowledge in audiovisual performance from a range of practitioners into a generic software toolkit. AV performers are specialists in sound visualization and visual sonification – audiovisual crossmodal interactions. We believe that their AV design skills were an important contribution to this research that benefited the design of the ofxAVUI technology. Additionally, these users make high demands for an AV system in terms of media manipulation and interaction design: they can be considered super-users, who are regularly performing in front of an audience, and need powerful, fluid and responsive manipulation of AV media through a robust interface.

We used hackathons as a motivating, productive way to connect with our users. Hackathons were employed from two different perspectives. In Stage 1, two hackathons were used for the rapid prototyping of AV performance systems by AV artists. In Stage 3, the hackathon aimed to test in a short period of time the ease of implementation of our toolkit. Participants were asked to develop a project with it, which could be added to our toolkit repository, as an extension of its functionality or a demonstration of a new use case. We also reached out to the community of interaction designers and developers following our GitHub repository website. In this sense, we complemented the "local" perspective of the hackathons with the "global" community of GitHub users, adapting, albeit on a smaller scale, the approach followed in [28].

Multi-stage Approach

Informed by related multi-stage studies [33, 35], we adopted a three-stage approach to the development of the ofxAVUI toolkit. This could be used for the development of other technologies. It can be summarized as *explore-consolidate-abstract* approach: 1) explore and gather multiple views via participatory design process, and evaluate results with other users; 2) design a prototype that consolidates best practices detected in the previous study, and evaluate with users; 3) develop a general technology based on the evaluation of the prototype, convert and abstract positive aspects of it into a toolkit, and run an additional participatory design and evaluation session for testing. This approach, with a participatory stage based on hackathons, allowed us to iteratively develop both our AVUI Guidelines and the ofxAVUI toolkit. Conclusions from each study were converted into design specifications for the following one. Conclusions from each stage fed into design specifications for the following one, with the last stage informing a set of general design guidelines.

AVUI as Parallel to Crossmodality in the Real World

Our interactions with the world are multi-sensorial. Opening a door handles produces auditory and visual feedback. Some of these interactions, such as pouring water into a glass, give us audiovisual feedback regarding dimensional data. In these interactions, audio and visual information are related in a congruent way. The concept of auditory icons aims to "to use sound in a way that is analogous to the use of visual icons to provide information," providing a a natural way to represent dimensional data as well as conceptual objects [10]. AVUI extends Gaver's pragmatic concept by proposing a crossmodal approach that incorporates UI elements, sonic feedback and congruent visualization in a way that aesthetic content and interface become one. The integrated audio and visualization reflect the status of UI elements, recalling a functional simplicity of the sort encouraged by John Maeda [22].

CONCLUSION

We have introduced the concept of AudioVisual User Interface (AVUI), a type of interface where UI, audio and visualization are interconnected and integrated. By combining UI with interrelated sound and image, the proposed concept of AVUI (and ofxAVUI toolkit in particular) can help leverage the identified benefits of audiovisual crossmodal interaction, such as improvements in performance and engagement.

We presented an iterative multi-stage process of design, prototyping, development and evaluation of ofxAVUI, an enabling software toolkit to facilitate development of AVUIs. The toolkit has been released as open-source, and is multi-platform, aiming to facilitate its adoption. Participatory design methods were used, centered around three hackathons. This process also allowed us to incorporate expert and practitioner insight into a series of generic guidelines for the design of AVUIs. The toolkit and guidelines will be of interest to interaction designers who wish to create compelling products and systems which integrate sound and image in the interface. By extending Hook's existing theoretical framework to study a large number of AV systems, we believe that the findings have a strong generalizability that the previous studies do not.

We believe that the AVUI concept and the ofxAVUI technology have potential for application in a number of use cases where a screen and interaction is shown to an audience, and for and multimodal interaction. The crossmodal linkages that an AVUI facilitate could be useful for engagement in VR and AR interactive environments. This form of interaction which fuses sensing modalities, function and content, can be compelling for a number of domains: not only areas where engagement is important, such as art, education and games, but also assistive and accessible technologies.

ACKNOWLEDGEMENTS

We would like to thank all who collaborated in this research: workshop and hackathon participants; evaluators and interviewees; colleagues; and in particular Borut Kumperščak for programming assistance and Alessandro Altavilla for documentation. This work was supported by the EU Marie Curie fellowship FP7 REA grant 627922 and also partly by FCT/MCTES LARSyS (UID/EEA/50009/2013 (2015-2017)).

REFERENCES

1. Paul Bach-y Rita. 1987. Brain Plasticity as a Basis of Sensory Substitution. *Neurorehabilitation and Neural Repair* 1, 2 (June 1987), 67–71. DOI: http://dx.doi.org/10.1177/136140968700100202

2. Florent Berthaut, Mark T. Marshall, Sriram Subramanian, and Martin Hachet. 2013. Rouages: Revealing the Mechanisms of Digital Musical Instruments to the Audience. In *Proceedings of New Interfaces for Musical Expression (NIME)*. Daejeon, Seoul. http://nime.org/proceedings/2013/nime2013_51.pdf

3. Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. DOI: http://dx.doi.org/10.1191/1478088706qp063oa

4. Ana Carvalho and Cornelia Lund (Eds.). 2015. *The Audiovisual Breakthrough*. Fluctuating Images, Berlin. http://www.ephemeral-expanded.net/audiovisualbreakthrough/

5. Karen Collins and Peter J. Taillon. 2012. Visualized sound effect icons for improved multimedia accessibility: A pilot study. *Entertainment Computing* 3, 1 (Jan. 2012), 11–17. DOI: http://dx.doi.org/10.1016/j.entcom.2011.09.002

6. Nuno N. Correia. 2013. *Interactive Audiovisual Objects*. Doctoral thesis. Aalto ARTS Books, Helsinki. https://www.taik.fi/kirjakauppa/product_info.php?cPath=26&products_id=269&language=en

7. Nuno N. Correia and Atau Tanaka. 2014. User-Centered Design of a Tool for Interactive Computer-Generated Audiovisuals. In *Proc. 2nd ICLI*. Lisbon. http://www.liveinterfaces.org/proceedings2014.html

8. Bram Crevits. 2006. The Roots of VJing. In *VJ: Audio-Visual Art and VJ Culture*, Michael Faulkner (Ed.). Laurence King Publishing, London, 14–19.

9. Carl DiSalvo and Melissa Gregg. 2013. Issue-Oriented Hackathons. (2013). http://carldisalvo.com/posts/issue-oriented-hackathons/

10. William W. Gaver. 1986. Auditory Icons: Using Sound in Computer Interfaces. *Hum.-Comput. Interact.* 2, 2 (June 1986), 167–177. DOI: http://dx.doi.org/10.1207/s15327051hci0202_3

11. F. Wai-ling Ho-Ching, Jennifer Mankoff, and James A. Landay. 2003. Can you see what i hear?: the design and evaluation of a peripheral sound display for the deaf. In *Proc. CHI'03 (CHI '03)*. ACM, New York, NY, USA, 161–168. DOI:http://dx.doi.org/10.1145/642611.642641

12. Lars Erik Holmquist. 2008. Bootlegging: Multidisciplinary Brainstorming with Cut-ups. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008 (PDC '08)*. Indiana University, Indianapolis, IN, USA, 158–161. http://dl.acm.org/citation.cfm?id=1795234.1795260

13. Jonathan Hook, David Green, John McCarthy, Stuart Taylor, Peter Wright, and Patrick Olivier. 2011. A VJ centered exploration of expressive interaction. In *Proc. CHI'11 (CHI '11)*. ACM, New York, NY, USA, 1265–1274. DOI: http://dx.doi.org/10.1145/1978942.1979130

14. Jonathan Hook, John McCarthy, Peter Wright, and Patrick Olivier. 2013. Waves: exploring idiographic design for live performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2969–2978. DOI: http://dx.doi.org/10.1145/2470654.2481412

15. Michael Xuelin Huang, Will W. W. Tang, Kenneth W. K. Lo, C. K. Lau, Grace Ngai, and Stephen Chan. 2012. MelodicBrush: A Novel System for Cross-modal Digital Art Creation Linking Calligraphy and Music. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 418–427. DOI: http://dx.doi.org/10.1145/2317956.2318018

16. Lilly Irani. 2015. Hackathons and the Making of Entrepreneurial Citizenship. *Science, Technology & Human Values* (April 2015), 1–26. DOI: http://dx.doi.org/10.1177/0162243915578486

17. Sergi Jordà. 2003. Sonigraphical Instruments From FMOL to the reacTable*. In *NIME03*. Montreal. http://www.nime.org/proceedings/2003/nime2003_070.pdf

18. M. Komssi, D. Pichlis, M. Raatikainen, K. Kindstrom, and J. Jarvinen. 2015. What are Hackathons for? *IEEE Software* 32, 5 (Sept. 2015), 60–67. DOI: http://dx.doi.org/10.1109/MS.2014.78

19. Michael Kubovy and Michael Schutz. 2010. Audio-Visual Objects. *Review of Philosophy and Psychology* 1, 1 (March 2010), 41–61. DOI: http://dx.doi.org/10.1007/s13164-009-0004-5

20. Golan Levin. 2000. *Painterly Interfaces for Audiovisual Performance*. Ph.D. Dissertation. M.Sc. Dissertation. Massachusetts Institute of Technology, Cambridge, MA. http://acg.media.mit.edu/people/golan/thesis/thesis300.pdf

21. Thomas James Lodato and Carl DiSalvo. 2016. Issue-oriented hackathons as material participation. *New Media & Society* (Feb. 2016), 1461444816629467. DOI: http://dx.doi.org/10.1177/1461444816629467

22. John Maeda. 2006. *The Laws of Simplicity*. The MIT Press, Cambridge, MA.

23. Thor Magnusson. 2006. Affordances and Constraints in Screen-based Musical Instruments. In *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles (NordiCHI '06)*. ACM, New York, NY, USA, 441–444. DOI: http://dx.doi.org/10.1145/1182475.1182532

24. Harry Mcgurk and John Macdonald. 1976. Hearing lips and seeing voices. *Nature* 264, 5588 (Dec. 1976), 746–748. DOI:http://dx.doi.org/10.1038/264746a0

25. Oussama Metatla, Nuno N. Correia, Fiore Martin, Nick Bryan-Kinns, and Tony Stockman. 2016. Tap the ShapeTones: Exploring the Effects of Crossmodal Congruence in an Audio-Visual Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1055–1066. DOI: http://dx.doi.org/10.1145/2858036.2858456

26. William Moritz. 1997a. Digital Harmony: The Life of John Whitney, Computer Animation Pioneer. 2.5 (Aug. 1997). http://www.awn.com/mag/issue2.5/2.5pages/2.5moritzwhitney.html

27. William Moritz. 1997b. The Dream of Color Music, And Machines That Made it Possible. *Animation World Magazine* 2.1 (April 1997). http://www.awn.com/mag/issue2.1/articles/moritz2.1.html

28. Alistair Morrison, Donald McMillan, Stuart Reeves, Scott Sherwood, and Matthew Chalmers. 2012. A Hybrid Mass Participation Approach to Mobile Software Trials. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1311–1320. DOI: http://dx.doi.org/10.1145/2207676.2208588

29. Paul Nagle. 2008. Yamaha Tenori-On Review. *Sound On Sound* Feb. 2008 (2008). http://www.soundonsound.com/sos/feb08/articles/yamahatenorion.htm

30. I. Olowe, G. Moro, and M. Barthet. 2016. residUUm: user mapping and performance strategies for multilayered live audiovisual generation. In *Proceedings of International Conference on New Interfaces for Musical Expression (NIME)*. Brisbane. http://www.nime.org/proceedings/2016/nime2016_paper0053.pdf

31. Shin'ichi Onimaru, Taro Uraoka, Naoyuki Matsuzaki, and Michiteru Kitazaki. 2008. Cross-modal information display to improve driving performance. In *Proc. ACM symposium on Virtual reality software and technology (VRST '08)*. ACM, New York, NY, USA, 281–282. DOI: http://dx.doi.org/10.1145/1450579.1450653

32. Chris Salter. 2010. *Entangled: Technology and the Transformation of Performance*. MIT Press, Massachusetts.

33. Guy Schofield, David Green, Thomas Smith, Peter Wright, and Patrick Olivier. 2014. Cinejack: Using Live Music to Control Narrative Visuals. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 209–218. DOI: http://dx.doi.org/10.1145/2598510.2598520

34. Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. 2000. Illusions: What you see is what you hear. *Nature* 408, 6814 (Dec. 2000), 788–788. DOI: http://dx.doi.org/10.1038/35048669

35. Alexander Wiethoff and Sven Gehring. 2012. Designing Interaction with Media Façades: A Case Study. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 308–317. DOI: http://dx.doi.org/10.1145/2317956.2318004

The Role of Live Visuals in Audience Understanding of Electronic Music Performances

Nuno N. Correia
Madeira-ITI, U. Madeira
Funchal, Portugal
nuno.correia@staff.uma.pt

Deborah Castro
Madeira-ITI, U. Madeira
Funchal, Portugal
deborah.castro@m-iti.org

Atau Tanaka
Goldsmiths, U. London
London, UK
a.tanaka@gold.ac.uk

ABSTRACT

There is an identified lack of visual feedback in electronic music performances. Live visuals have been used to fill in this gap. However, there is a scarcity of studies that analyze the effectiveness of live visuals in conveying feedback. In this paper, we aim to study the contribution of live visuals to the understanding of electronic music performances, from the perspective of the audience. We present related work in the fields of audience studies in performing arts, electronic music and audiovisuals. For this purpose, we organized two live events, where 10 audiovisual performances took place. We used questionnaires to conduct an audience study in these events. Results point to a better audience understanding in two of the four design patterns we used as analytical framework. In our discussion, we suggest best practices for the design of audiovisual performance systems that can lead to improved audience understanding.

CCS CONCEPTS

• **Human-centered computing → Graphical user interfaces**; *Information visualization*; • **Applied computing → Performing arts**;

KEYWORDS

Audiovisuals, live visuals, music, performance, audience

ACM Reference Format:
Nuno N. Correia, Deborah Castro, and Atau Tanaka. 2017. The Role of Live Visuals in Audience Understanding of Electronic Music Performances. In *Proceedings of AM '17, London, United Kingdom, August 23–26, 2017,* 8 pages.
https://doi.org/10.1145/3123514.3123555

AM '17, August 23–26, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5373-1/17/08…$15.00
https://doi.org/10.1145/3123514.3123555

1 INTRODUCTION

Laptops have become a common tool in electronic music performances, with the appearance of powerful portable computers capable of real-time audio manipulation. However, the introduction of the laptop on the stage also brought with it an absence of visual feedback and gestural information regarding the performer's actions to the audience [3, 21]. Amongst the approaches artists have taken is to use live visuals in electronic music performances to compensate for this absence [12].

We can find roots to today's approaches in live visuals in the experiments of the early-mid 20th century, with artists such as Oskar Fischinger combining elaborate film projection setups with music [25]. In turn, these influenced a generation of artists who would collaborate with live music artists in the 1960s and 1970s, with psychadelia-influenced approaches such as liquid projections [26]. The popularization of powerful personal computers with multimedia capabilities in the 1990s empowered a new generation of visual and audiovisual performers to generate and manipulate digital content in real time, live [33]. Audiovisual (AV) artists often create their own systems for performance, with a DIY approach, using software and programming environments such as Max/MSP, PureData, Processing and openFrameworks [11]. Salter describes that this practice adopted "a long litany of names such as audiovisual performance, real-time video, live cinema, performance cinema, and VJ culture" [33].

AV artists have aimed to combine sound and image for different reasons. One of these is to achieve a "total artwork" or "gesamtkunstwerk", using a "common urgence of every art towards the most direct appeal to a common public" [35], or in other words a "hypermediacy" [5]. Another approach has been to overcome the above-mentioned problem of lack of visual feedback in electronic music performances, by conveying visually to an audience, via video projection, all the "elements implied on the process of making their own music" [21], or in other words contributing to "immediacy" [5].

Practitioners have pursued two main paths to visualization: 1) using graphical correspondences more or less tightly mapped to sound and interaction parameters, defined as audiovisual performance, an "interconnection between sound and image, which sometimes becomes apparent and at other times remains intuitive" [6], and 2) projecting to the audience the code generated by music performers who use programming techniques to produce sound, a practice known as "live coding" [8]. In this paper, we will focus on the former. These experiences might carry risks of obscuring this understanding, not clarifying it: "visuals are an overpowering medium and can easily detract performance" [22]. Moreover, they created new issues regarding audience understanding of the processes and interaction behind the visuals.

Most studies that focus on the lack of visual feedback in laptop music and audiovisual performances (such as [21]) have approached this problem from the perspective of the performer. We propose to address this from the perspective of the audience, to answer the following research question: "What design approaches in laptop-based audiovisual performances can be more conducive to audience understanding of the performer's actions?" To answer it, we have developed a study involving the participatory design and development of a series of audiovisual projects, and their presentation in two public performance events, where audience studies were conducted.

2 RELATED WORK

Audience Research

The audience, one of the constituents of communication process, has been traditionally considered a passive or, even, an "invisible" and "mute" agent [36]. Semiotics and media studies started to explore a possibility of an active role. For example, interpretative media theory is based on the assumption that media texts are polysemic and open to individuals' interpretations [15]. Likewise, the concept of *open text* coined by Eco [14] refers to the multiple readings a single text may embrace beyond the one envisaged by the author. However, the asymmetry (or lack of correspondence) between the codes used by the reader and the author may lead to distorted communication [19].

Different methods have been used to understand spectators' responses to cultural objects (e.g. movies, novels, plays, songs). The nature of these objects affects the relationship between them and the audience. For instance, a reading audience usually engages with a stable text in privacy, while a theater audience not only interacts with a "transitional cultural object" but can also be part of a "collective effervescence" [17]. The nature of the cultural object will, therefore, determine the type of audience analysis carried out.

In the performing arts, different approaches to audience studies have been pursued. The research project "Watching Dance: Kinesthetic Empathy" [31] combined both qualitative audience methods and neurophysiological research to analyze how spectators respond to, and identify with, dance during and after the performance. Albert [1] analyzed how dancers and audience members react to choreographed and improvised movements in social dance by using conversation and video analysis. In their exploration of the tele-presence and performing arts convergence, Cesar et al. [7] used galvanic skin response sensors to analyze the engagement of theatre remote audiences, while Radbourne et al. [29] conducted focus groups to measure quality in theatrical works and live music concerts with an emphasis on potential reattendance.

In music, Jaimovich et al. [20] designed an installation to collect large samples of physiological and self-reported data using questionnaires to analyze people's emotional reaction to recorded music. In order to explore the response of young adults to a chamber music concert, Dearn and Pitts [13] relied on the combination of questionnaires, Write-Draw cards and focus groups. Lai and Bovermann [23] carried out semi-structured interviews with participants of a live aural performance with electronic instruments "to understand the communication flow and the engagement between performer and audience". In this paper, we employed questionnaires to analyze the perceived correlation, from the audience's point of view, between the actions of the performer and visuals – the perceived transparency of the performance.

Fels [16] defines *transparency* as "the psychophysiological distance, in the minds of the player and the audience, between the input and the output of a device mapping". On the basis that music is a multisensory phenomenon [34], visual kinematic information from an event (e.g. performer actions) has been documented to be a crucial factor in the communication of meaning [28] and in the emotional reactions evoked [34] within the audience. Nonetheless, electronic music performances are often sensor or laptop-based, which are not always visible to the public and whose usage does not require big gestures and actions from the performer. The configuration of electronic music performances transmits to the spectator little or, even, no-information about what is happening on the stage, unlike performances with acoustic instruments [3]. This may result in an unclear cognitive link between the sound and the performer's actions or, in the terminology of Fels et al. [16], in a lack of transparency.

In electronic musical instruments, the decoupling of control and sound makes transparent mapping a challenge [16]. Transparency can be explored from two different points of view: that of the player and that of the audience. On the one hand, "transparency of a mapping for the player depends both on cognitive understanding and on physical proficiency". On the other hand, the audience may need "to have an understanding of how the instrument works to appreciate the proficiency of the player" which may be affected by, for example, their own cultural knowledge [16]. Interest and enjoyment levels may be independent of the audience's technical knowledge. Indeed, Bin et al. [4] noted that having access to tutorials on digital musical instruments before the performance increased the audience's know-how on the topic but did not seem to have an impact on these quantities.

Gurevich and Fyans [18] pose the following questions regarding the spectator's experience: 1) Address: How does the spectator know that the performer is interacting with the system?, 2) Attention: How does the spectator know that the system is responding to the performer?, 3) Action: How does the spectator think the user controls the system?, 4) Alignment: How does the spectator know that the system is doing the right thing?, 5) Accident: How does the spectator know when the performer of the system has made a mistake? In particular, we focus our attention on "action" – called "mapping comprehension" by Barbosa et al. [2] – that digs into audience's understanding of how the mapping between input (cause) and output (effects) functions. In other words: "How clear is the relationship between the performer's actions and, in our particular case, the audiovisual result?".

Taxonomy for Audiovisual Systems and AVUI

There are a large variety of real-time audiovisual art approaches that have been used in electronic music performances. Ribas [32] has created a taxonomy that classifies these different systems, based on four categories:

(1) *Audiovisual entities*, assuming "distinct procedural behaviors and responses to interaction": pieces composed of distinct individual elements, "mostly graphic shapes or moving pixels", which have "associated sound excerpts or loops, either to graphic forms or to an overall visual configuration".
(2) *Interactive sounding shapes*, where specific audiovisual elements are not necessarily created through the user's interaction, "but rather chosen, selected, altered, added or activated – reconfigured within the possibilities given by an existing repertoire devised within the system".
(3) *Sounding figurations*, consisting of visual elements that can be drawn or created by "screen-based and mouse-operated systems" and "whose properties are mapped to the production of synthetic sounds". In this category, nothing happens without human interaction, since "it is exactly human expression that the system is devised to integrate and express as its subject matter", producing "consistent responses to user input". This relates to Levin's work and research on painterly interfaces for audiovisual performance [24].
(4) *Audiovisual reactions to interactions*, where "changes to the audiovisual surface are a response to the participants' combined actions", often gestural. Reactions to the behavior are indeterminable: "there can be no linear correspondence" between an interaction and an audiovisual reaction "due to the fluctuating nature of the input data".

Looking at Ribas's taxonomy [32] from the perspective of audience understanding, the categories of *audiovisual entities* and *sounding figurations* seem more conductive to audience understanding. The former, because discreet user interactions are mapped to each entity, allowing for a more analytical representation. The latter, because of its inherent tight and consistent mapping between drawing and audiovisual result. These two categories relate to our concept of Audiovisual User Interface (AVUI) [11], where "UI, audio and visualization are interconnected and integrated", leading to clear audiovisual responses to interaction.

Interactive sounding shapes and *audiovisual reactions* seem to be on the other side of the spectrum. The former, because they are not necessarily created through the performer's interaction. The latter, because of the absence of linear mapping between interaction and reaction. To validate and further develop the nascent AVUI concept, we conducted the hackathons as a way to get audience feedback on the prototypes made by our artists. This led to broader insights about the role of visuals in the audience understanding of electronic music performance, which we report here.

3 METHODS

Preliminary Stage: Interviews

In previous research, we adopted a participatory design approach for the development and study of software for AV performances. We conducted interviews with 12 audiovisual performers, and asked them about their practice, the creative tools they use, and also their needs and desires as performers. One of the key themes identified in the analysis of the interviews was related to the communication of the performance process to the audience, and audience understanding [9].

Audiovisual Projects

The topic of audience understanding, and other key themes detected, informed a sketching workshop, and two hackathons (Gen.AV 1 and 2), on the topic of creating new tools for performance with generative audiovisuals. Five projects were created in Gen.AV 1, and six in Gen.AV 2. In [11], we have reported on these hackathons, and subsequent evaluation of projects. This evaluation consisted of tests by other performers of the projects, over a period of around one week, followed by interviews to gather feedback. The procedures followed were:

- Hackathon >Audience Study >Peer Evaluation

These procedures were repeated twice, in succession (once for Gen.AV 1 and another for Gen.AV2). We will now present the relevant projects resulting from this process, where user

Figure 1: Image from first event, Modulant project.

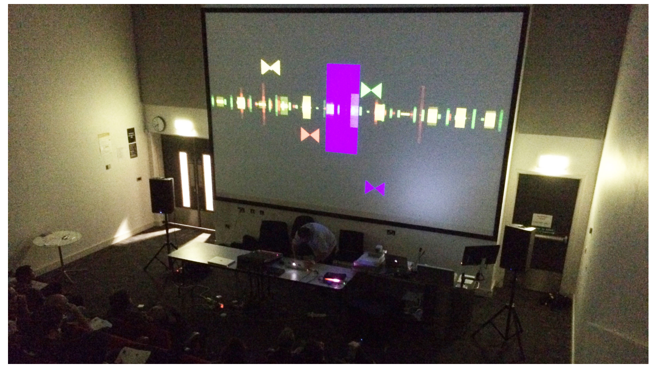


Figure 2: Image from second event, Butterfly project.

interface elements were often integrated into sound visualizations, creating *audiovisual entities*, or AVUI.

Four projects were presented for audience evaluation at the Gen.AV 1 event. ABP is an animation engine and sound visualizer where the user can define visual parameters using an overall GUI (Graphical User Interface). Esoterion Universe starts with an empty 3D space that can be filled with planet-like audiovisual objects. GS.avi is an instrument that generates spatial visualizations and music from the gestures of a performer. Modulant allows for drawing images, which are then sonified. Six projects were presented at Gen.AV 2 event. Butterfly allows for the visualization and control of four audio synthesizers, by manipulating four icons in the screen. Cantor Dust generates, displays, and sonifies cantor set fractals. EUG (Esoterion Universe Gestenkrach) further develops Esoterion Universe from Gen.AV 1, adding 3D gestural control. OnTheTap plays with the tactile, analog feel of tapping surfaces as interaction input, combined with a GUI. residUUm allows for the creation of audiovisual particles with a specific lifespan, volume and panning [27]. Wat creates a chaotic audiovisual texture based on cellular automata, distributed in a rectangular 3D space. More information about all Gen.AV projects: (http://www.gen-av.org) and [10, 11].

These 10 projects can be mapped to Ribas's taxonomy for audiovisual pieces [32]:

(1) Butterfly, Esoterion University, EUG and residUUm can be considered *audiovisual entities*, as they are composed of individual elements, with distinctive graphic shapes and associated sounds, and also a related user-interface.
(2) ABP, Cantor Dust and OnTheTap can be classified as *interactive sounding shapes*, as they do not intended to represent specific sounds, but consist of generic graphics reacting to the overall sonic landscape, with the possibility of reconfiguring both audio and visuals by means of an interface independent from the visual output.
(3) Modulant belongs to the *sounding figurations* category, as it consists of graphical elements that are drawn and mapped to the production of synthetic sound.
(4) GS.avi and Wat can be considered as part of the *audiovisual reactions to interactions* category, as changes to the audiovisual surface are a response to the performer's gestural actions, in a non-linear correspondence.

Audience Studies

In order to study audience experience in audiovisual performances, we organized two public events, one per hackathon, at Goldsmiths, University of London, in February (Figure 1) and July 2015 (Figure 2). Each project was presented in a ten-minute performance, followed by a short discussion by the authors (around five minutes). In these events, we distributed questionnaires to the audience, with questions targeting each of the projects, four for Gen.AV 1 and six for Gen.AV 2 (one of the five Gen.AV 1 projects was left out of questionnaires due to late completion and last minute addition to event). In Gen.AV 1, 45 respondents answered the questionnaire, and 34 respondents in Gen.AV 2. The audience filled in the questionnaires in the short intervals between project performances.

The questionnaires consisted of three pairs of questions, repeated according to the number of projects (four for Gen.AV 1 and six for Gen.AV 2). Each pair consisted of a 5-point likert scale, and an open-ended question. Two of the pairs asked concern variety/diversity of audio and visual content, and relatedness between both modalities:

1) Did you find that the audiovisuals were varied and diverse? Rate: (1-5)
Complete the sentence: The audio and visuals were...
2) Did you find that sounds and visuals were

well related? Rate: (1-5)
Complete the sentence: The relationship between sounds and visuals was...

These two pairs of questions are outside the scope of this paper. A third pair of our questions related to audience understanding of the performer's actions:

3) Did you find the connection between the performer's actions and the audiovisual result understandable? Rate: (1-5)
Complete the sentence: The performance was...

This was informed by Gurevich and Fyans's "audience questions" presented above, namely the "action" one [18]: "How does the spectator think the user controls the system?". In our formulation, members were asked to rate 1-5, from 1-"not understandable at all", to 5-"very understandable". They were also asked to provide further insight on the topic in an open-ended question, by requesting them to complete the sentence "The performance was...".

4 RESULTS

45 audience members that answered the questionnaire in the Gen.AV 1 event, with an average age of 29. In terms of gender, 27 identified themselves as male, 12 as female, and 6 did not indicate gender. From the Gen.AV 1 audience members, 69% had experience as practitioner in audio and/or visuals: 27% had experience as visual artist, visual designer or VJ, 18% as musician or DJ, and 24% in both. Age average of the 34 respondents to the Gen.AV 2 questionnaire was 33. In this case, 19 were male, 14 female, and 1 did not fill in this section. From the Gen.AV 2 audience members, 79% had experience as practitioners: 26% as visual artist, visual designer or VJ, 38% as musician or DJ, and 15% in both. The event was promoted at the location of the event, Goldsmiths, University of London, a university with a strong arts and music tradition, and at the London Music and Video Hackspace communities. Therefore the profile of the audience fits with our expectations – an audience composed of members familiar with diverse music and audiovisual performance practices.

Figure 3 presents the summary of the results obtained from the question on audience understanding of performers actions (Likert scale) as a boxplot. Projects Esoterion Universe and Modulant from Gen.AV 1, and Butterfly, EUG and residUUm from Gen.AV 2 obtained the highest results (Modulant with a median of 5, the others with a median of 4). As we presented in the Methods section, these are the projects corresponding to the *audiovisual entities* and *sounding figurations* categories. This confirms our hypothesis that projects in these categories have higher potential for audience understanding.

The *sounding figurations* approach of Modulant was particularly successful: "Interesting and clear, it was quite easy

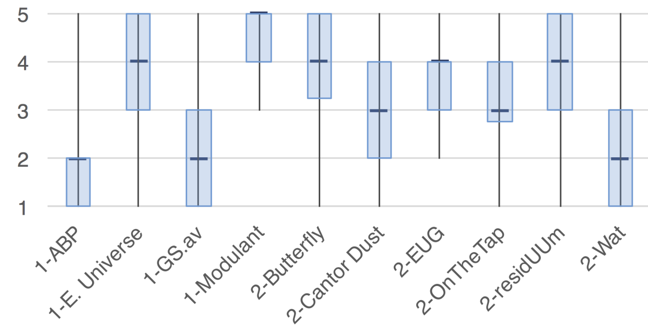


Figure 3: Boxplot with results from audience understanding of performers actions in Gen.AV 1 and 2 projects (5-point Likert scale in y-axis).

to understand how user input affects sound". A respondent found it clearer than the "conventional" live coding: "the next layer for live coders to add to their tool kit, such that the audience get more of an understanding of the process being instigated on stage, rather than seeing just code (only select audiences can read such code)".

The UI quadrants of Butterfly, with its XY parametric space and four "butterfly" icons representing the current state, seems to have been also been clear to the audience, as confirmed by the following observations: "largely comprehensible via point and click style GUI"; "moving 'bow-ties' [butterflies] to control the music"; "this was clear by seeing the cursor moving the individual butterflies"; "cursor movement in XY space with animated bars made for clean interaction"; and "really clear".

residUUm was praised for its stylized cursor: "a nice cursor, much better than other performances' quotidian cursor", pointing in the same direction of some of the expert observations regarding the stylistic integration of the GUI with the visuals. Despite the success of the cursor in residUUm in attracting attention and conveying interaction, not all aspects of mechanics behind the software were made clear, raising further questions: "The cursor was a focal point when moved the visuals changed. Maybe click and hold cursor makes it louder?". The issue of UI stylization (or lack of) is also implicit regarding OnTheTap: "Clear but basic in terms of its interaction. It would have been nice to see a more elegant solution rather than sliders".

Some contrasting comments highlighted more negative aspects regarding visible interaction. While some respondents enjoyed seeing the cursor and the controls, others find viewing a cursor on the screen unappealing: "not nice to see a mouse cursor in a live performance" (Esoterion Universe). Also, some respondents found one project (the highest scoring one in this regard, Modulant) too clear in its interaction:

"Perhaps too much [understandable]. Not very fluid, but synchronized. Simple."; "didactic"; and "like a demo".

Not all understanding of the performance was conveyed by the visuals. When information was lacking on the screen, the audience appeared to want to read what the performer was doing from his/her body, as illustrated by the following statements: "It was not possible to see their actions. Maybe use some cameras next time" (AVP); "Not clear because I could not see the performer" (GS.avi); "hard to understand as I could not tell what was being done by the performer" (Wat). In some cases, the understanding only occurred during the post-performance explanation: "I didn't understand what was going on until I was told" (GS.avi); "when I heard the explanation about tap it was interesting but didn't see it in performance" (OnTheTap).

5 DISCUSSION

Visibility of interface and of the parameter space – the interval of values a parameter can have, and its current position on that interval – seem to have contributed to audience understanding. Four of the five projects that achieved the highest scores in the questionnaire (medians of 4 and 5) made visible both the interface and the parameter space. One of these five projects, residUUm, did not show a graphical interface, but highlighted the agency of the performer with a stylized cursor, and showed the parameter space. Cantor Dust and OnTheTap, both with a median of 3, implemented only one of these aspects (visibility of parameter space, in Cantor Dust) or only temporarily showed them (OnTheTap). The remaining projects, with a median of 2, employed neither.

We will now identify design solutions which have led to better results with audience members. Our outcomes point to a higher audience understanding within the categories, or design patterns, of *audiovisual entities* and *sounding figurations*. We firstly analyze elements within each of the categories that offer potential for audience understanding.

Regarding *audiovisual entities*, four aspects have been explicitly mentioned by audience members: 1) the visibility of the interface per entity or module; 2) the stylization of the GUI; 3) the parameter space of that interface; 4) the legibility of performer's actions through the cursor. All four *audiovisual entities* projects followed these aspects (with the exception of residUUm, which did not implement aspect 1).

Concerning *sounding figurations*, the open-ended comments of audience members regarding Modulant mention important aspects to take into account for these types of projects. Although understanding in Modulant was very high, both as reflected in its score and in open-ended comments, not all respondents were pleased with the transparency and direct mapping between drawing and sound. This raises the issue if, after a certain threshold, a high understanding of the performance might be detrimental to the experience, making it appear less like an act of expression and more as a technical demonstration. As recommended by Levin when discussing painterly interfaces, it might be beneficial to "eschew mappings based on the arbitrary conventions of visual language, or the arbitrary affordances of computational technologies", and pursue more dynamic mappings, using gesture and correspondent animation properties such as velocity, orientation and curvature [24].

The analysis of the results allows us to propose best practices for the design of live visuals leading to better audience understanding in performances:

- Adopt a design pattern based on *audiovisual entities* or *sounding figurations*. Both patterns create a direct link between interaction and result. In the case of *audiovisual entities*, this connection is apparent by creating multiple 'objects' consisting of corresponding audio and visual elements, which can be controlled independently. In the case of *sounding figurations*, the connection comes from consistent mapping strategies between the act of drawing or the resulting figure, and a sonic result.
- For the *audiovisual entities* design pattern, ensure that: 1) each audiovisual module has a visible corresponding interface (and not an overall UI or control panel, as usually happens in AV software); 2) the graphic design of the interface matches the aesthetic of the visuals, to create a coherent whole; 3) the parameter space is visible for each module (minimum, maximum and current status of the parameters); 4) the agency of the performer is present on the screen, but in a stylized way to avoid a "demo effect" (for example, with a customized cursor or symbolic representation of multitouch). Recommendations 1-3, on the whole, match the concept of AVUI [11].
- Regarding a *sounding figurations* design pattern, do not employ a simplistic approach, which might appear over overly demonstrative. This can be accomplished by avoiding direct one-to-one mappings between points on the screen and audio properties, and introducing dynamic elements (speed, orientation etc) from the gesture generating the drawing.

We have analyzed the importance of visuals to communicate interaction information to the audience. But there are other factors in audiovisual performance, external to the visuals, that contribute to audience understanding. We identify three: 1) the body of the performer; 2) the eventual explanation by the performer; and 3) the characteristics of the setting and the audience itself. Even in laptop-based performances, the audience looks for visual cues from the body of the performer. We found that this was particularly the case when the visuals did not assist in communicating the performer's actions.

Another element that influences audience understanding is the artist explanation. In our events, we asked performers to explain their pieces to the audience, after their performance. Our results suggest that these explanations influence audience understanding as well. However, we believe that it did not overly influence the results for this study, as audience members distinguished between meaning acquired from explanation and from the performance itself.

One last external element is the setting for the performance, and the profile of the audience itself. These seem to us related, as from our experience, a certain setting tends to attract a determined profile, forming a loose community. For example, in our case the performances took place in Goldsmiths, University of London, a university well known for its art and musical studies. That would lead to a more knowledgeable audience regarding performative practices, electronic music and audiovisuals. Our experience also tells us that there is a specialized audience for laptop electronic music and audiovisual performance, particularly in larger urban environments. More research should be conducted on characterizing audiences for audiovisual performances.

6 CONCLUSIONS

We have addressed the issue of audience understanding in laptop-based music performances with live visuals. We analyzed this topic, from the perspective of audience members, in two events (10 different audiovisual projects presented in total). Our results confirmed the polysemic nature intrinsic to the interpretative process of cultural objects.

We have presented best practices in the design of software for audiovisual performance leading to audience understanding, regarding two identified successful design patterns – *audiovisual entity* and *sounding figurations* – which relate to our concept of AVUI. We have also identified additional factors that influence audience understanding in laptop-based performance, and that deserve attention for future work: body of performer; performance explanation; and community (audience-setting).

Even though questionnaires are one of the most common methods used to understand audience response to an event [7], a more multi-layered approach would have allowed us to dig deeper in the process of audience understanding. In addition, gathering more information about the profile of the attendees would have helped in the interpretation of the data collected. For example, by exploring further their level of familiarity with the cultural object, their pre-formed taste, their prior judgment and, even, the expectation with which they arrive to the event. More data could be gathered in the future with a more diverse audience. Future research could aim at studying more examples of each project category, particularly within *sounding figurations* (only one example, Modulant, was included in this category). The projects presented are merely case-studies within each category, and their representativeness could be reinforced in the future with more examples.

We believe that more research needs to be carried out within the topic of live visuals and audience understanding, especially due to the identified limitations of laptop-based performance in terms of visual feedback, and the potential of live visuals to address that. Additional questions to be asked are: Does the live coding mode also satisfy audience communication? And: Does electronic music benefit from visuals, or are there other approaches conductive for audience understanding – such as augmenting 'traditional' instruments, or more emphasis on gestural control? Lastly: How does understanding relate to the overall audience enjoyment? Another aspect that deserves attention is how the reported experience of the audience might influence the work of the performers, or even, those members of the audience who are also practitioners. This idea not only aligns with the outcomes of the present paper, focused on "understanding", but also emphasizes the complexity of the concept of "experience". According to Reason [30], "experience" is not merely "what is going on in an audience's mind (and body) during a performance, but also is what they do with this experience after the event". This particular perspective on the audience has been mostly neglected, even though it is here "where there is rich potential for developing strong, creative and self-reflective methodological approaches" [30]. This provides with a path for future work – to adopt a broader perspective of audience experience, during and after the event.

There is a growing interest in audience participation in performances, and we maintain that understanding is a crucial condition for successful participation. Live visuals can be more than "blinking lights" [21] or an overpowering element [22] – they have the potential to successfully augment the electronic music performer, contributing to audience understanding.

7 ACKNOWLEDGEMENTS

We would like to thank all who collaborated in this research – hackathon participants, performers and audience members. This work was supported by the EU Marie Curie fellowship FP7 REA grant 627922 and also partly by FCT/MCTES LARSyS (UID/EEA/50009/2013 (2015-2017)) and MITI/Excell (M1420-01-0145-FEDER-000002-MITIExcell).

REFERENCES

[1] Saul Albert. 2015. Rhythmical coordination of performers and audience in partner dance. Delineating improvised and choreographed interaction. *Etnografia e ricerca qualitativa* 3/2015 (2015). https://doi.org/10.3240/81723

[2] Jerônimo Barbosa, Filipe Calegario, Veronica Teichrieb, Geber Ramalho, and Patrick McGlynn. 2012. Considering Audience's View Towards an Evaluation Methodology for Digital Musical Instruments. In *Proc. NIME 2012*. Ann Arbor. http://www.nime.org/proceedings/2012/nime2012_174.pdf

[3] Florent Berthaut, Mark T. Marshall, Sriram Subramanian, and Martin Hachet. 2013. Rouages: Revealing the Mechanisms of Digital Musical Instruments to the Audience. In *Proc. NIME 2013*. Daejeon, Seoul. http://nime.org/proceedings/2013/nime2013_51.pdf

[4] S. Astrid Bin, Nick Bryan-Kinns, and Andrew P. Mcpherson. 2016. Skip the Pre-Concert Demo: How Technical Familiarity and Musical Style Affect Audience Response. In *Proc. NIME 2016*. 200–205. http://www.nime.org/proceedings/2016/nime2016_paper0041.pdf

[5] Jay David Bolter and Richard Grusin. 2000. *Remediation: Understanding New Media.* The MIT Press, Massachusetts.

[6] Ana Carvalho and Cornelia Lund (Eds.). 2015. *The Audiovisual Breakthrough.* Fluctuating Images, Berlin. http://www.ephemeral-expanded.net/audiovisualbreakthrough/

[7] Pablo Cesar, Jack Jansen, Erik Geelhoed, Doug Williams, Ian Kegel, Marian F. Ursu, and Chen Wang. 2015. Distributed Theatre: Connecting (with) Remote Audiences. In *Proceedings of the Workshop on Everyday Telepresence: Emerging Practices and Future Research Directions, in conjuncion with the ACM SIGCHI Conference on Human Factors in Computing Systems*. Seoul.

[8] Nick Collins, Alex McLean, Julian Rohrhuber, and Adrian Ward. 2003. Live Coding in Laptop Performance. *Org. Sound* 8, 3 (Dec. 2003), 321–330. https://doi.org/10.1017/S135577180300030X

[9] Nuno N. Correia and Atau Tanaka. 2014. User-Centered Design of a Tool for Interactive Computer-Generated Audiovisuals. In *Proc. 2nd ICLI*. Lisbon. http://www.liveinterfaces.org/proceedings2014.html

[10] Nuno N. Correia and Atau Tanaka. 2015. Prototyping Audiovisual Performance Tools: A Hackathon Approach. In *Proc. NIME 2015*. Baton Rouge, Louisiana. http://www.nime.org/proceedings/2015/nime2015_230.pdf

[11] Nuno N. Correia and Atau Tanaka. 2017. AVUI: Designing a Toolkit for Audiovisual Interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1093–1104. https://doi.org/10.1145/3025453.3026042

[12] Bram Crevits. 2006. The Roots of VJing. In *VJ: Audio-Visual Art and VJ Culture*, Michael Faulkner (Ed.). Laurence King Publishing, London, 14–19.

[13] Lucy K. Dearn and Stephanie E. Pitts. 2017. (Un) popular music and young audiences: Exploring the classical chamber music concert from the perspective of young adult listeners. *Journal of Popular Music Education* 1, 1 (2017), 43–62. https://doi.org/10.1386/jpme.1.1.43

[14] Umberto Eco. 1993. *Lector in Fabula* (2nd edition ed.). Fabbri, Milano.

[15] William A. Evans. 1990. The interpretative turn in media research: innovation, iteration, or illusion? *Critical Studies in Mass Communication* 7 (1990), 147–168.

[16] Sidney Fels, Ashley Gadd, and Axel Mulder. 2002. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound* 7, 02 (2002), 109–126. https://doi.org/10.1017/S1355771802002042

[17] Wendy Griswold, Elizabeth Lenaghan, and Michelle Naffziger. 2011. Readers as Audiences. In *The Handbook of Media Audiences*, Virginia Nightingale (Ed.). Wiley-Blackwell, 17–40. http://onlinelibrary.wiley.com/doi/10.1002/9781444340525.ch1/summary DOI: 10.1002/9781444340525.ch1.

[18] Michael Gurevich and A. Cavan Fyans. 2011. Digital Musical Interactions: Performer-system relationships and their perception by spectators. *Organised Sound* 16, 2 (2011), 166–175. https://doi.org/10.1017/S1355771811000112

[19] Stuart Hall. 1999. Encoding, Decoding. In *The Cultural Studies Reader*. 507–517.

[20] Javier Jaimovich, Miguel Ortiz, Niall Coghlan, and R Benjamin Knapp. 2012. The Emotion in Motion Experiment: Using an Interactive Installation as a Means for Understanding Emotional Response to Music. In *Proc. NIME 2012*. 457–458. http://www.nime.org/proceedings/2012/nime2012_254.pdf

[21] Vitor Joaquim and Álvaro Barbosa. 2013. Are Luminous Devices Helping Musicians to Produce Better Aural Results, or Just Helping Audiences Not To Get Bored?. In *xCoAx 2013: Proceedings of the first conference on Computation, Communication, Aesthetics and X.* http://2013.xcoax.org/#p36

[22] Sergi Jordà. 2005. *Digital Lutherie - Crafting Musical Computers for New Musics' Performance and Improvisation.* PhD thesis. UPF, Barcelona. http://mtg.upf.edu/node/449

[23] Chi-Hsia Lai and Till Bovermann. 2013. Audience Experience in Sound Performance. In *Proc. NIME 2013*. 170–173. http://nime.org/proceedings/2013/nime2013_197.pdf

[24] Golan Levin. 2000. *Painterly Interfaces for Audiovisual Performance.* Ph.D. Dissertation. M.Sc. Dissertation. Massachusetts Institute of Technology, Cambridge, MA. http://acg.media.mit.edu/people/golan/thesis/thesis300.pdf

[25] William Moritz. 2004. *Optical Poetry: The Life and Work of Oskar Fischinger.* John Libbey Publishing, Eastleigh.

[26] Glenn O'Brien. 2005. Tune In, Turn On, Light Up... *Tate Etc.* 4 (2005). http://www.tate.org.uk/tateetc/issue4/summeroflove.htm

[27] Ireti Olowe, Giulio Moro, and Mathieu Barthet. 2016. residUUm: user mapping and performance strategies for multilayered live audiovisual generation. In *Proc. NIME 2016*. Brisbane. http://www.nime.org/proceedings/2016/nime2016_paper0053.pdf

[28] Friedrich Platz and Reinhard Kopiez. 2012. When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception* 30, 1 (2012), 71–83. https://doi.org/10.1525/mp.2012.30.1.71

[29] Jennifer Radbourne, Katya Johanson, Hilary Glow, and Tabitha White. 2009. The Audience Experience: Measuring Quality in the Performing Arts. *International Journal of Arts Management* 11, 3 (2009), 16–29. http://www.jstor.org/stable/41064995http://about.jstor.org/terms

[30] Matthew Reason. 2010. Asking the audience: audience research and the experience of theatre. *About performance* 10 (2010), 15–34. https://ray.yorksj.ac.uk/id/eprint/381/

[31] Dee Reynolds. 2011. Watching Dance: Kinesthetic Empathy. (2011). http://www.watchingdance.org/

[32] Luísa Ribas. 2011. *The Nature of Sound-Image Relations in Digital Interactive Systems.* PhD thesis. Universidade do Porto Faculdade de Belas Artes, Porto.

[33] Chris Salter. 2010. *Entangled: Technology and the Transformation of Performance.* MIT Press, Massachusetts.

[34] Jonna K. Vuoskoski, Marc R. Thompson, Charles Spence, and Eric F. Clarke. 2016. Interaction of Sight and Sound in the Perception and Experience of Musical Performance. *Music Perception: An Interdisciplinary Journal* 33, 4 (April 2016), 457–471. https://doi.org/10.1525/mp.2016.33.4.457

[35] Richard Wagner. 2001. Outlines of the Artwork of the Future. In *Multimedia: From Wagner to Virtual Reality*, Randall Packer and Ken Jordan (Eds.). W. W. Norton, New York, 3–9.

[36] Brita B. Wheeler. 2004. The social construction of an art field: how audience informed the institutionalization of performance art. *The Journal of Arts Management, Law, and Society* 33, 4 (2004), 336–350. https://doi.org/10.3200/JAML.33.4.336-350

Tap the ShapeTones: Exploring the Effects of Crossmodal Congruence in an Audio-Visual Interface

Oussama Metatla
Queen Mary University of London
o.metatla@qmul.ac.uk

Nuno N. Correia
Goldsmiths, University of London
n.correia@gold.ac.uk

Fiore Martin
Queen Mary University of London
f.martin@qmul.ac.uk

Nick Bryan-Kinns
Queen Mary University of London
n.bryan-kinns@qmul.ac.uk

Tony Stockman
Queen Mary University of London
t.stockman@qmul.ac.uk

ABSTRACT
There is growing interest in the application of crossmodal perception to interface design. However, most research has focused on task performance measures and often ignored user experience and engagement. We present an examination of crossmodal congruence in terms of performance and engagement in the context of a memory task of audio, visual, and audio-visual stimuli. Participants in a first study showed improved performance when using a visual congruent mapping that was cancelled by the addition of audio to the baseline conditions, and a subjective preference for the audio-visual stimulus that was not reflected in the objective data. Based on these findings, we designed an audio-visual memory game to examine the effects of crossmodal congruence on user experience and engagement. Results showed higher engagement levels with congruent displays with some reported preference for potential challenge and enjoyment that an incongruent display may support, particularly for increased task complexity.

Author Keywords
Crossmodal congruence, spatial mappings, user engagement, user experience, games, audio-visual display.

ACM Classification Keywords
H.1.2 User/Machine systems: Human Factors; H.5.2 User Interfaces: Auditory (non-speech) feedback, Screen design (e.g., text, graphics, color), Input devices and strategies (e.g., mouse, touchscreen), Evaluation/methodology; K.8.0 General: Games

INTRODUCTION
Multisensory perception is an activity that we do everyday when we combine signals from various sensory channels to make sense of our environment and to act in it. One of the mechanisms that we use to fuse input from multiple sensory channels is referred to as *crossmodal interaction* [35]. A key feature of a crossmodal display is that it relays the same information through two or more senses, for example, when we find it easier to recognise speech when we can see the speaker's lip movements. Research aiming to apply findings from crossmodal perception to interface design has focused on designing support for interaction in complex environments, for example in the design of monitoring systems and warning signals [29, 35], on designing sensory substitution devices for people with sensory disabilities, such as the vOICe system, which uses sonification to convert images into sound [15], and on supporting collaboration between people with different sensory abilities [42, 16]. However, whilst it is increasingly feasible to support crossmodal interaction in a range of general purpose devices, e.g. tablet computers and smartphones provide touch, visual, and speech interaction, little work has considered the implications of crossmodal displays on user experience and engagement. Therefore, we propose that research into the design of effective crossmodal interfaces should consider a wider range of user experiences. In particular, evaluations of crossmodal displays should emphasise elements of both user performance and engagement to provide deeper insights into the application of crossmodal mappings to interactive experiences. This paper contributes to bridging the gap between studies of crossmodal user performance and engagement by examining the effects of crossmodal congruence on performance and engagement in the context of a memory task supported by combinations of audio and visual displays on touch-screen devices. A first study examines the effects of different levels of crossmodal congruence on how audio-visual cues support the mapping of spatial ordering. A second study examines the application of these crossmodal mappings in the design of an audio-visual memory game, focusing on evaluating user experience and engagement with the crossmodal gameplay.

BACKGROUND
Crossmodal interaction underlies the phenomenon by which signals from one sensory modality can affect the process-

ing of information perceived through another modality. One famous example of this phenomenon is the "McGurk" effect [14] where the auditory phoneme "ba" is perceived as "da" when paired with the visual stimuli of lips movements pronouncing "ga". The ideas behind crossmodal interaction stem from advances in cognitive neuroscience, specifically new understandings of brain plasticity and sensory substitution, which refer to the capacity of the brain to replace the functions of a given sense by another sensory modality [1]. Recently, there has been increasing interest in the study of these types of crossmodal interactions between sensory information, and their implications for user interface design. For instance, Ju-Hwan and Spence [12] demonstrated that the presentation of sounds can modulate the number of vibrotactile targets that a person will perceive, particularly when they perform secondary attention-demanding tasks. Shams *et al.* [31] also demonstrated how people's perception of flashing lights can be manipulated by sounds, with people seeing a single flash of light as consisting of two flashes when these are presented simultaneously with multiple auditory beeps. Sensory modalities are therefore far from working as independent modules and findings from these and similar studies challenge the notion that their interaction follows a hierarchy in which vision dominates the sensory experience. Massaro [13] suggests that while all modalities contribute to perceptual experience, it is most influenced by the sensory channel that mediates the least ambiguous information. In the context of this paper, this suggests that different visual and auditory mappings can influence the perception of spatial information and that different combinations may result in more efficient and engaging interactions.

Congruency and crossmodal correspondences

Research examining multi-sensory experience often use the term *congruence* or *crossmodal correspondences* to refer to non-arbitrary associations that exist between different modalities and the consequences that these have on human information processing. For instance, studies found crossmodal correspondences between high-pitched sounds and bright, small objects positioned at higher locations in space, and between low-pitched sounds and darker, bigger rounder objects at lower locations [2, 26]. Other studies found congruent mappings between pitch and vertical location, size and spatial frequency [5]. Spence highlights a further distinction between *semantic* and *synaesthetic congruency* to differentiate between sensory stimuli that vary in terms of their identity and/or meaning, and those that refer to *"correspondences between putatively nonredundant stimulus attributes or dimensions that happen to be shared by many people"* [34]. A number of researchers have demonstrated the benefits of exploiting crossmodal congruency for better user interface design. Hoggan and Brewster [9], for instance, examined the relationships between individual visual button features such as size and height with audio/tactile properties. They showed that perceived quality of touchscreen buttons was correlated to congruence between visual and audio/tactile feedback used to represent them. Fewer researchers have looked at user experience - Huang *et al* developed the MelodicBrush system in which they explored how crossmodal mappings between the shapes of Chinese calligraphy and musical tones can enhance user experience during artistic creation [10], but their system did not ground mapping choices in empirical data.

Attention, memory & motor learning

To explore crossmodal interaction we are interested in how semantically congruent audio and visual stimuli can convey spatial information and guide users' attention when locating items on interactive touch-screen devices as there is extensive evidence supporting the existence of crossmodal links in spatial attention (for reviews, see [33, 36]). In particular, a number of lab-based studies have demonstrated how the presentation of crossmodal as opposed to unimodal cues can significantly facilitate the capture of a person's spatial attention [37]. Stefanucci and Proffitt [38] examined the impact of crossmodal cues on memory tasks involving visual and auditory stimuli with a focus on whether congruency effects between learning and retrieval phases improves retention. Their findings indicated that the presence of sounds provided a strong cue for binding visual display to the information learnt and hence improve retention. Studies have also shown that crossmodal concurrent feedback can enhance motor learning, and positive effects are often explained by a reduction of workload [7]. For example, visual feedback could facilitate learning of spatial aspects of the movement, while auditory feedback could support learning of temporal aspects [32]. Curiously, concurrent crossmodal feedback has been found to enhance performance in the acquisition phase, but the performance gains are lost in retention tests. This finding is explained by the guidance hypothesis which states that permanent feedback during acquisition leads to a dependency on the feedback [30]. The guidance forces learners to ignore their intrinsic feedback which is based on proprioception [32] - our sense of bodily movement and position in space.

SCOPE

The work presented in this paper extends this line of research by examining more complex crossmodal stimuli that support a spatial ordering memory task and by exploring crossmodal congruences from two perspectives: i) task performance, and ii) user engagement. Study 1 builds on previous work on crossmodal perception that demonstrated congruence effects between the auditory feature of pitch and the visual features of size and vertical location [2, 26, 5], as well as the impact of crossmodal feedback on motor learning and the retention of spatial information post-acquisition [7, 30]. The aim is to evaluate users' ability to determine spatial orderings of a sequence of items on the basis of audio, visual, and audio-visual stimuli, in the context of a memory task. Study 2 explores user experience and engagement with a crossmodal memory game building on the results of the first study, and recent work on the evaluation of user engagement in game applications [23, 24, 41]. To do this, we added a number of gamification elements to the apparatus used in the first study that were inspired by current design practices in mobile games. These included the introduction of a game progression logic based on increasing levels of difficulty and scores with corresponding visual and auditory indicators [28]. This is described in more details in later sections of the paper.

STUDY 1: CROSSMODAL MAPPINGS

Apparatus

To examine the impact of crossmodal congruence on mappings of vertical location, motor learning and retention, we designed an interface that consists of a visual and an auditory display component for output and a touch-based component for input. The experimental apparatus was developed as an application that runs on an Apple iPad. It divides the screen horizontally into different sections (top of Figure 1), with each section corresponding to a unique shape and a unique tone (bottom of Figure 1) we refer to as *ShapeTones*.

Visual mappings & congruence levels

We designed three types of visuals to map screen sections; we refer to these as *arbitrary*, *size*, and *spikes* (Figure 1). In the spikes mapping, we used a basic circular shape and increased the amount of spikes attached to it to correspond to a given section; e.g. the shape for section three has three spikes. There is therefore an immediately perceivable relationship between the shapes and the physical layout of the screen, which constitutes a *congruent* mapping. In the size mapping, we used a single shape and we varied its size to correspond to each screen section. The gradual change in size therefore corresponded to the progression of sections, with lower sections corresponding to larger objects [2, 26]. However, compared to the spikes mapping, the exact mapping from a given size to a section has to be inferred. This mapping is therefore *semi-congruent* with the physical layout. In the arbitrary shapes mapping, different shapes are assigned arbitrarily to correspond to each screen section. We designed these shapes so that they bear no obvious relationship to the sections and are therefore *incongruent* with the physical layout of the screen.

Auditory mapping

Tones were mapped vertically to screen sections: lower pitches to lower sections, and higher pitches to higher sections (Figure 1). This mapping is based on crossmodal correspondences between vertical location and pitch [2, 26, 5]. We used musical notes and the sine wave timbre as they are common tones. After trying different scales in terms of tone discernibility with iPad speakers, we chose a mid-range octave: the G4 major scale.

Touch-based input

Users interact with this application by tapping on corresponding sections on the screen to reproduce the spatial order of a sequence of items conveyed to them through ShapeTones.

Experimental design

We manipulated level of congruency as an independent variable in a between-subjects experimental design. Participants were divided into three groups with each group performing the experimental task using one of the three visual mappings; participants used the *spikes mapping* in the *congruent condition*; the *size mapping* in the *semi-congruent condition*; and the *arbitrary mapping* in the *incongruent condition*.

We also manipulated display type in a within-subjects experimental design. Participants in each group performed the experimental task under three within-subjects conditions; an *audio-visual condition*; *a visual-only condition*; and an *audio-only condition*. The audio-only and visual-only conditions were used as controls to provide baselines to compare crossmodal and unimodal displays, i.e. to examine the effects of the visual and auditory mappings when used independently as a means for judging locations on the touch-screen. A between-subjects design thus ensured that each participant is only exposed to one visual mapping/congruence level, while a within-subjects design ensured that each participant's performance with a given crossmodal congruence level is compared against their own performance on unimodal displays. The combination of between/within-subjects designs also avoids confounding learning effects and fatigue.

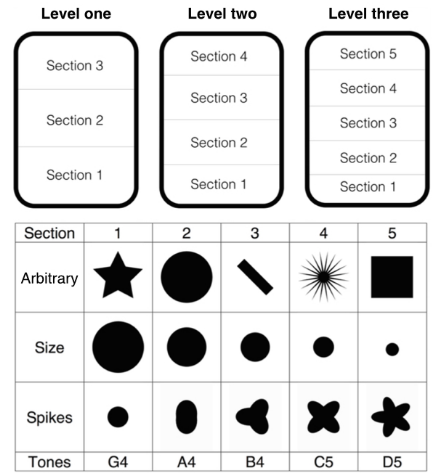


Figure 1. Crossmodal mappings.

Experimental task details

The experimental task was a memory task in which participants were presented with a sequence of three ShapeTones and were asked to reproduce the order of that sequence by tapping the corresponding sections on the touch-screen. This task builds on previous work in the area of point estimation [17] and provides a potential for broader use, e.g. in games. ShapeTones were presented one at a time at the centre of the touch-screen at a speed of 0.3 seconds per item chosen on the basis of previous studies on rapid identification of auditory and graphical stimuli [18, 27]. Depending on the experimental condition, participants were asked to watch and/or listen to a sequence of three shapes and tones and to reproduce the order in which these occurred by tapping on corresponding sections on the touch-screen.

Figure 2 exemplifies the structure of the experimental task. Participants tapped on a "play" button to start the sequence, watched and/or listened to a sequence, then tapped on the touch-screen to reproduce its order. No feedback was presented while tapping the order of the sequence, but the participants' input was played back to them at the end of the tapping (in the form of ShapeTones in the audio-visual condition, shapes only in the visual-only condition, and tones only in the audio-only condition). This was then followed by an in-

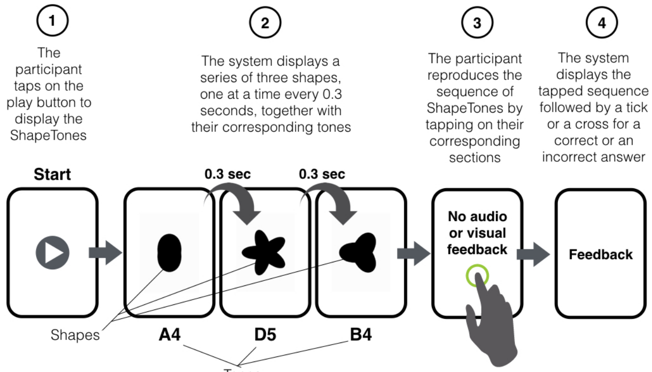


Figure 2. Experimental task

dication of whether their sequence was correct or not (a tick for a correct sequence, and a cross for an incorrect sequence). To avoid ceiling effects in each condition, participants performed sets of experimental tasks using three different complexity levels as shown in Figure 1. We used, three, four and five sections in each level of complexity respectively. The stimuli consisted of three ShapeTones in all complexity levels. Each participant performed 10 trials in each set, totalling 30 trials per condition; thus giving 90 trials per participant and a total of 3240 trials for the whole study.

Experimental setup & procedure

Participants were briefed about the study, signed consent forms and completed an initial questionnaire about demographic details, their musical training (in terms of years of practice), and experience with touch-screen devices. They were then randomly assigned to one of the three groups. Care was taken to ensure different musical abilities were broadly distributed between groups. Before the trials began, participants were trained on the particular display they were going to use. Unlike the tests, training was such that participants could tap around the touch-screen and receive audio, visual or audio-visual feedback that corresponded to the location of where they tapped. They were instructed to take as much time as they needed to memorise the tones and the shapes used for the particular condition they were about to do. Training typically lasted up to 5 minutes. Once familiar with the display, participants performed three trials similar to the actual the testing phase (Figure 2). These were not included in the analysis and were intended to help participants develop their proprioceptive skills. Participants then performed ten trials in each of the three conditions (audio-only, visual-only, and audio-visual) in a given level of complexity before moving on to the next level. They were allowed to familiarise themselves again with the shapes and tones before the start of each new set of trials. We administered short questionnaires and conducted informal interviews at the end of each level to collect feedback. Conditions were counterbalanced. An entire session lasted between 45 minutes and an hour.

Participants

36 participants took part in this study (19 female, 17 male, *mean age* $= 27.9, SD = 4.4$). They were a mixture of university staff (academic and non-academic), undergraduate and postgraduate students and members of the public. Participants received a cash incentive for participating. Seven participants rated their musical experience as expert, eight as intermediate, fourteen as beginner, and six had no musical training. All had experience with using touch-screen devices.

Dependent variables & measurements

The dependent variables were the scores and completion times. Scores were calculated based on the number of correct sequences reproduced by the participants. Completion times were measured as the duration from the time the participants pressed the "play" button to the instant they tapped the third and final point in a given sequence.

Hypotheses

S1H1: Level of congruence will have an effect on participants' performance: in particular, based on existing literature on crossmodal mappings [2, 5, 9, 26, 29], we expected a congruent display using the spikes mapping to lead to better performances than a semi-congruent display using the size mapping and an incongruent display using the shapes mapping. We also expected the semi-congruent display to yield better performances than the incongruent display.

S1H2: Type of display will have an effect on participants' performance: in particular, based on existing literature on the advantages of audio-visual over unimodal displays [4, 38], we expected that the effects of the level of congruence will be more apparent in the audio-visual conditions.

Results

We used single-factor ANOVAs with level of congruence as a factor (three levels: congruent, semi-congruent, and incongruent) to analyse differences in times and scores across groups, and repeated-measures ANOVAs with display type as a factor (three level: audio-visual, audio-only, and visual only) to analyse differences within each group. In both cases, we used Fisher's LSD for post-hoc comparisons of main effects. We used a confidence level of $\alpha = 0.05$ for all tests.

Scores across groups

Level one (three sections)

There was no significant main effect of level of congruence on participants' scores in the audio-visual ($F(2, 34) = 0.104, p = 0.902$) visual-only ($F(2, 34) = 1.578, p = 0.222$) and audio-only conditions ($F(2, 34) = 0.54, p = 0.588$).

Level two (four sections)

There was no significant main effect of level of congruence on participants' scores across groups in the audio-visual condition ($F(2, 34) = 2.834, p = 0.074$). In the visual-only condition, there was a significant main effect of level of congruence on scores ($F(2, 34) = 4.276 p = 0.023, \eta^2 = 0.21$). Post-hoc tests showed that participants in the congruent condition (spikes: $mean = 7.16, sd = 1.99$) scored significantly higher than participants in the semi-congruent condition (size: $mean = 5.33, sd = 1.37$) ($p = 0.015$). Participants in the incongruent condition (random: $mean = 7.18, sd =$

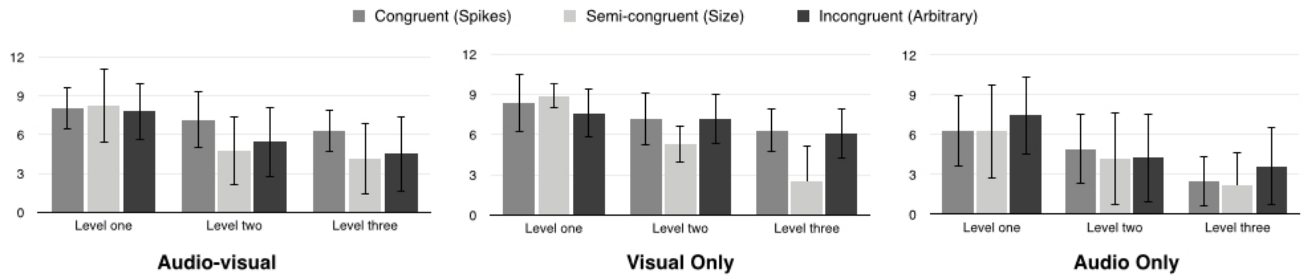


Figure 3. Scores across groups and complexity levels in the audio-visual, visual-only and audio-only conditions

1.88) also scored significantly higher than those in the semi-congruent condition ($p = 0.013$). There was no significant difference between the congruent and the incongruent conditions ($p = 0.984$). There was also no significant main effect of levels of congruence on participants' scores in the audio-only condition ($F(2, 34) = 0.192, p = 0.826$).

Level three (five sections)
There was no significant main effect of level of congruence on participants' scores across groups in the audio-visual condition ($F(2, 34) = 2.565, p = 0.093$). In the visual-only condition, there was a significant main effect of level of congruence on participants' scores across groups ($F(2, 34) = 12.097, p < 0.001$, $\eta^2 = 0.43$) and post-hoc tests showed that participants in the congruent condition (spike: $m = 6.33, sd = 1.66$) scored significantly higher than participants in the semi-congruent condition (size: $mean = 2.5, sd = 2.67$) ($p < 0.001$), and that participants in the incongruent condition (random: $mean = 6.09, sd = 1.86$) scored significantly higher than participants in the semi-congruent condition ($p = 0.001$). There was no significant difference between participants scores in the congruent and incongruent conditions ($p = 0.746$). There was no significant main effect of level of congruence on participants's scores across group in the audio-only condition ($F(2, 32) = 1.082, p = 0.351$).

Figure 3 summarises these results, which show that: as complexity increased, participants who used the congruent display (spikes mapping) performed significantly better than those who used the semi-congruent and incongruent displays in the visual-only conditions; Participants who used the incongruent display performed significantly better than those who used the semi-congruent display in the visual-only conditions; Augmenting the baseline visual mappings with audio output in the audio-visual conditions seems to have eliminated the observed effects of congruency levels.

Task completion times across groups

Level one (three sections)
There was no significant main effect of level of congruence on task completion times in the audio-visual ($F(2, 34) = 0.456, p = 0.638$), visual-only ($F(2, 34) = 0.679, p = 0.514$), and audio-only conditions ($F(2, 34) = 0.496, p = 0.614$).

Level two (four sections)
There was no significant main effect of level of congruence on task completion times across groups in the audio-visual condition ($F(2, 34) = 0.436p = 0.65$). In the visual-only condition, there was a significant main effect of level of congruence on task completion times across groups ($F(2, 34) = 3.72, p = 0.035, \eta^2 = 0.18$) and post-hoc tests showed that participants in the incongruent condition (random: $mean = 4649.1ms, sd = 971.6ms$) spend significantly longer time to complete the task than those in the semi-congruent condition (size: $mean = 3655.5ms, sd = 1179.2ms$) ($p = 0.04$) and those in the congruent condition (spikes: $mean = 3458.6ms, sd = 1162.8ms$) ($p = 0.015$). There was no significant difference between task completion times in the congruent and semi-congruent conditions ($p = 0.667$). There was also no significant main effect of level of congruence on task completion times across group in the audio-only condition ($F(2, 34) = 0.303, p = 0.74$).

Level three (five sections)
There was no significant main effect of level of congruence on task completion times across groups in any of the audio-visual ($F(2, 34) = 2.977, p = 0.065$), visual-only ($F(2, 34) = 2.792, p = 0.076$) and audio-only conditions ($F(2, 32) = 1.805, p = 0.181$). The above results showed that participants who used the congruent and semi-congruent display were significantly faster than those who used the incongruent display, but this was the case only in the visual-only condition in complexity level two (where the screen was divided into four sections). Again, the introduction of audio output in the audio-visual condition seems to have eliminated these significant effects.

Results within each group

We combined data from all levels of complexity to analyse scores and task completion times within each group.

Congruent group
There was a significant main effect of display type on participants' scores in the congruent group ($F(2, 22) = 13.307, p = 0.001, \eta^2 = 0.547$). Post-hoc tests showed that participants in this group scored significantly higher in the visual-only condition ($mean = 21.41, sd = 4.99$) compared to the audio-only condition ($mean = 13.75, sd = 6.48$) ($p = 0.004$). Their scores in the audio-visual condition ($mean = 21.58, sd = 4.88$) were also significantly higher than in the audio-only condition ($p = 0.002$). Differences between their scores in the audio-visual and visual-only conditions were not statistically significant ($p = 0.861$).

There was also a significant main effect of display type on participants' task completion times ($F(2,22) = 16.584, p = 0.001, \eta^2 = 0.601$). Post-hoc tests showed that participants spent significantly longer times to complete the task in the audio-only condition ($mean = 5574.5, sd = 2100.2$) compared to the visual-only condition ($mean = 3537.5, sd = 1250.9$) ($p = 0.001$), and to the audio-visual condition ($mean = 4325.7, sd = 1898.3$) ($p < 0.001$). There was no statistically significant difference between task completion times in the visual-only and audio-visual conditions ($p = 0.071$). The above results show that participants performance was best when using a visual-only display and that the combination of audio-visual output in a congruent display increased performance times without significantly improving scores.

Semi-congruent group

The effect of display type on participants in the semi-congruent group was not significant for scores ($F(2,22) = 2.216, p = 0.133$) but it was significant for task completion times ($F(2,22) = 3.369, p = 0.043, \eta^2 = 0.249$). For the latter, post-hoc tests showed that participants spent significantly longer times to complete the task in the audio-only condition ($mean = 4579.9, sd = 1608.6$) compared to the visual-only condition ($mean = 3670, sd = 1002.1$) ($p = 0.042$). Difference between the visual-only condition and the audio-visual condition ($mean = 4456.8, sd = 1507.6$) were also statistically significant with participants spending longer time in the audio-visual condition ($p = 0.009$). There was no statistically significant difference in task completion times between the audio-only and the audio-visual conditions ($p = 0.779$). These results show that combining auditory and visual output in a semi-congruent display increased performance times and levelled the scores across the three types of displays.

Incongruent group

There was a significant main effect of display type on participants' scores in the incongruent group ($F(2,22) = 6.212, p = 0.013, \eta^2 = 0.383$). Post-hoc tests showed that participants scored significantly higher in the visual-only condition ($mean = 17.81, sd = 5.61$) compared to the audio-only condition ($mean = 20.9, sd = 4.15$) ($p = 0.011$) and to the audio-visual condition ($mean = 15.36, sd = 8.15$) ($p = 0.024$). The differences in scores between the audio-only and the audio-visual conditions were not statistically significant ($p = 0.18$). There was no significant main effect of display type on task completion times in this group ($F(2,22) = 1.082, p = 0.358$). These results show that combining auditory and visual output in an incongruent display did not have a significant impact on scores and performance times.

Discussion

Our hypothesis that the congruent spikes mapping leads to better performances was only partially confirmed. Participants in the congruent group scored significantly higher than participants in the other groups but only when using a visual-only display. One of the most interesting findings in Study 1, which goes against our initial hypothesis is that the effects of the level of congruence seem to have been cancelled by the introduction of audio to the baseline visual conditions. Differences between participants' performances across groups in the audio-visual conditions were not statistically significant, which suggests participants relied on the audio output to compliment or compensate for the discrepancies in congruence levels used in the size and arbitrary mappings.

We note that a number of participants reported that they sometimes chose to ignore the shapes in the audio-visual conditions. This in turn suggests that those participants relied on the audio output as a primary source for determining spatial orderings of sequences of items, which should mean that they would perform well in the audio-only conditions. The objective data contradicts this analysis, however, showing performances in the audio-only conditions to be overall worse across all complexity levels. The shape mappings used in the audio-visual conditions supported better performances albeit at the expense of more effort.

But our hypothesis that participants would perform significantly better when using audio-visual as opposed to unimodal displays was also not fully supported since participants' scores across the three groups were consistently and often significantly higher in the visual-only conditions. These findings contrast those reported in the literature which often report advantages of crossmodal over unimodal cues in recognition and retention tasks [4, 38]. Interestingly, subjective feedback from the majority of participants did not reflect the analysis obtained from the objective data. For example, many participants across the three groups described how the speed of presentation of the shapes made the task more difficult to complete in the visual-only conditions and that the addition of tones improved this experience. In a recent study, Guastello et al [8] found that auditory and visual stimuli presented at intervals of about 300ms often produce miss errors in one or the other channel, which could explain the lower scores we obtained in the audio-visual conditions. A possible explanation for these seemingly contradictory accounts is that participants' answers in the interviews and questionnaires reflected perceived as opposed to actual difficulty. The addition of the tones to the crossmodal display may therefore have improved their confidence without necessarily impacting their scores.

Our expectation regarding the semi-congruent mapping was also not confirmed. We expected the semi-congruent size mapping to provide better support for remembering spatial locations than an incongruent arbitrary mapping, but our results showed this the opposite to be the case. The size mapping we used exploits previously reported crossmodal correspondences between vertical location, pitch and object size [2, 26, 5], but the type of task we used in our study could be a possible explanation for why these correspondences did not yield better performance. Whereas crossmodal correspondences have been studied almost exclusively in laboratory settings with simple cues where participants often deal with single or dual items [6, 33], our results show that retention of a sequence of multiple items appears to be more challenging and thus requires more careful design of crossmodal support. Indeed, as complexity increased, participants in the semi-congruent group highlighted that whilst they were able to identify that an extreme location had occurred in a sequence (i.e. small and large shapes), they found it increasingly challenging to accurately reproduce full sequences, particularly those including the middle ranges of the screen (sections two, three and four). So, we suggest that whilst requiring significantly more time to complete, the distinctive visual characteristics of the shapes used in the arbitrary mapping provided a better mapping in this case.

Interestingly, a number of participants from the incongruent group highlighted that whilst they found it challenging to focus on both the shapes and the tones in the audio-visual condition, they also felt that this challenge made the task more enjoyable and engaging. None of the participants in the other groups expressed this opinion when asked about their experiences and preferences. Thus, subjective feedback indicates that, although the incongruent display did not offer complimentary information, the challenge of combining incongruent information across auditory and visual modalities increased enjoyability and engagement with the task.

From the interviews we found that there were two distinct types of responses to the addition of tones to the crossmodal displays. The first was that tones were treated as a dominant output mode, with the shapes ignored or used as a secondary source of spatial information. This was often reported to be the case in the incongruent arbitrary mapping group. The second was that participants preferred to use the shapes as the dominant source of spatial information with tones used as a secondary channel. This was often the case in the congruent and semi-congruent displays. We also observed that participants tended to switch to this "complimentary strategy", where reliance on the secondary modality increased, as the task increased in complexity. These observations are inline with claims that crossmodal perception is most influenced by the sensory channel that mediates the least ambiguous information [13] and that the positive effects of crossmodal concurrent feedback can be explained by a reduction of workload [7]. Our results confirm these findings and highlight that levels of congruency can be a factor in determining complementarity of information display.

STUDY 2: USER ENGAGEMENT

Given the subjective feedback reported in Study 1, we ran a second study focusing on engagement and user experience. This complements the focus on performance-based measures in Study 1, and follows a trend within HCI studies to take experiential issues into account, emphasizing *"the experience of using the technology, rather than the focus on the task that is characteristic of many other approaches HCI"* [11] and aiming to understand *"how the user makes sense of the artefact and his/her interactions with it at emotional, sensual, and intellectual levels"* [43]. This trend has often been ignored in the study of crossmodal displays. In order to facilitate an experience that would be more conducive to engagement and enjoyment, and in response to the reported appeal of 'challenge' identified in Study 1 (particularly with arbitrary shapes), we adapted our test application into a game. With a few exceptions, digital games use the potential of visual display for aesthetic appeal and for elements of game design more than auditory display [20]. Among the exceptions in the field of mobile games are the *Papa Sangre* series

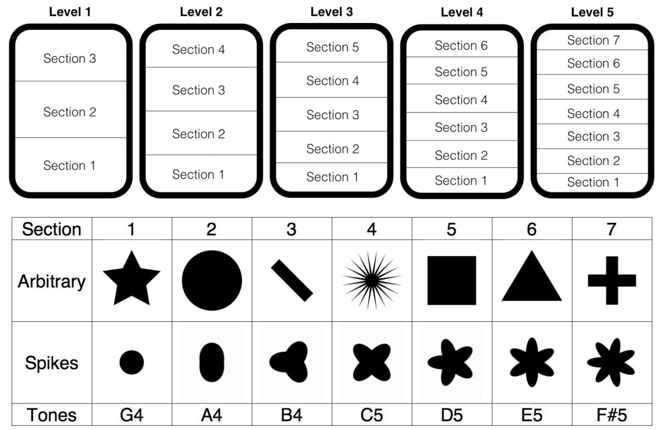


Figure 4. Levels and sections (top), crossmodal mappings (bottom)

[40] and *Dark Echo* [19], which are audio-focused games that cannot be played without sound. Nacke *et al* [20] have reported on the importance of auditory display for gameplay experience across different experiential dimensions (immersion, tension, competence, flow, negative and positive affect, and challenge). We were therefore also interested in examining the role of crossmodal display in gameplay.

Apparatus

We added a number of game design elements based on current design practices in games, particularly mobile games. Levels are a common concept – as the player succeeds in a task, she/he moves to a higher level, often with a higher degree of difficulty. We added a level identity and introduced a progression logic where the environment and challenge remains unchanged unless the player passes the challenge – in which case, the difficulty level will increase. The levels, sub-levels and the progression logic we added were as follows:

1. A player moves to a next sub-level upon successful completion of a trial. A new sequence of ShapeTones is then generated.

2. After ten sub-levels, a new level starts, with an additional ShapeTone – one vertical section is added to the initial three, and so forth, up to seven (Figure 4).

3. A training area is presented to the player at the beginning of each level for testing the new ShapeTones.

4. If the player fails a trial, the sub-level does not progress and the same ShapeTone sequence is played again.

Another common element in games is the score. The usage of scores and visual metaphors such as stars, to indicate degree of success or progress, are common gamification instruments [28]. We added a two-tier score – a star score (1-3 stars, dependent on performance on that level), and a numerical score. The scores and progression feedback we added were:

1. A trial score of one to three stars for passing a sub-level based on speed of playing back the correct sequence (one for slower, three for faster).

Figure 5. Game images, from left to right: arbitrary shapes version; spikes version; feedback after a trial; followed by global score.

2. Win and lose graphics and sounds for the end of each trial. (Figure 5).
3. The sub-level score accumulates in a global score, presented to the player at the end of each trial (Figure 5).

All other functionalities and application design remained unchanged from Study 1.

Study Design

We aimed to examine user engagement with two versions of the resulting game. We used the spikes and arbitrary shape mappings for each version because these emerged as the most successful visuals mapping in terms of performance in the first study. Several instruments have been developed to measure engagement in games, such as the Game Engagement and Game Experience Questionnaires [21]. Some of these instruments give particular attention to the concept of flow, such as the GameFlow model [39]. However, a range of diverse features contribute to engagement in games [3], we therefore adopted for an instrument that takes into account this diversity, the User Engagement Scale (UES) [24], more specifically the UESz version [41], which unlike other game engagement instruments, has been empirically validated. UES is a self-report measure consisting of a 31-item measured as a 5-point Likert scale that takes into account multiple dimensions of engagement: aesthetic appeal, perceived usability, felt involvement, novelty, focused attention and endurability. [24]. The UES has also been used in comparative studies [22]. Wiebe *et al.* [41] revised the UES for use in games (renamed as UESz) by organising the measures into four subscales: Focused Attention (FAz), Perceived Usability (PUz), Aesthetics (AEz) and Satisfaction (SAz). Later studies on the UES agree with the UESz revised set of subscales [22].

We used a within-subject design and invited participants to play the two versions of the game for 10 minutes each (10 minutes was the minimum duration of gameplay in similar studies deploying the UESz [41]). Before each 10 minute session, participants could play with the game for a short while (typically around 3 minutes), to get acquainted with it. The sequence of versions of the game was randomised and counterbalanced. We then asked participants to fill in UESz questionnaires for each version of the game and logged their scores for later analysis. An additional reason to use UESz as the sole questionnaire was to avoid respondent fatigue, as respondents already had to answer the UESz twice – one for each version of the game. Finally, we conducted a short interview focusing on crossmodal issues, usability and overall satisfaction with the game. We asked questions about the appeal of the tones and shapes, and how well they were connected. We also asked what was more important to play the game: tones, shapes, or both; and if it could be played with only audio or visuals. We also asked participants if they found something frustrating, and if they would play the game again.

Subscale	Arbitrary	Spikes	Significance
FAz	3.4 (0.77)	3.8 (0.85)	W = -2.4, $p = 0.016$
PUz	2.8 (0.81)	3.3 (0.78)	W = -2.764, $p = 0.006$
AEz	3.1 (0.62)	3.6 (0.63)	W = -2.371, $p = 0.018$
SAz	3.8 (0.75)	4.0 (0.84)	W = -1.963, $p = 0.05$

Table 1. *Mean* (*S D*) of UESz subscales for arbitrary shapes and spikes.

Hypothesis

S2H1: The congruent spikes mapping will be more engaging than the incongruent arbitrary shapes mapping.

Participants

Twelve participants took part in this study different from those who took part in the first study (10 male and two female, *mean age* = 36.6, *sd* = 6.7). Participants were a mixture of university staff and students. All participants received a cash incentive for participating. When asked about previous experience in games, on a scale of 1 (not at all experienced) to 5 (very experienced), only one participant declared to be very experienced, with two additional ones answering 4 (*mean* = 2.8). Most of the participants (seven) considered themselves to be very experienced as musicians, with two additional ones answering 4 (*mean* = 3.9). Only one participant considered himself to be very experienced as a visual artist, with three additional ones answering 4 (*mean* = 3.3).

Results

We used a Wilcoxon signed-rank test to analyse data from the UESz questionnaires and Student t-test to compare logged scores. Results from the questionnaire revealed a statistically significant preference for the spikes mapping in all four UESz subscales (Table 1). This preference was higher for Focused Attention (FAz), Perceived Usability (PUz) and Aesthetics (AEz) with differences of 0.4 to 0.5 between means, and less marked with Satisfaction (SAz), with a difference of 0.2 between means. Participants also performed better using the spikes mapping, reaching on average a higher levels (*max level mean* = 39.5, *sd* = 4.6) than with the arbitrary shapes mapping (*max level mean* = 33.1, *sd* = 8.1). This difference was statistically significant ($t = -2.579$, $p = 0.026$).

Based on interview responses, 10 participants preferred the spikes mapping and stated that the relationship between tones and shapes was more effective with spikes. Two participants did not find the two visual mappings to be very different, with one expressing a preference for arbitrary shapes because they were more distinguishable, and another stating that he did not identify any relationship between shapes and tones. One participant who preferred the spikes mapping stated that arbitrary shapes *"are more noticeable, but harder to concentrate* *[on]"*. Another participant considered that the spikes mapping became more difficult to distinguish in higher levels.

When asked what was more important to play the game, tones or shapes, six of the participants answered that they mostly relied on tones; four stated that they played it mostly as a visual game; one participant mentioned that he alternated between focusing on tones and shapes depending on the visualisation type; and another mentioned that he almost did not notice the visuals. Independently of the main modality, eight of the 12 participants stated that they would use the secondary modality as a backup when the difficulty level was higher. Sample statements, from visual-focused participants: *"when I got lost I relied on sound"*, *"I used sound as a backup"*, *"rely on image then rely on sound as a backup"*, *"sound was used as a check, as a support"*; and from sound-focused participants: *"I would get a visual as something to refer back"*, *"visual element gave me a confirmation"*, *"when you got the first one wrong and do it again, visuals become more important"*. Two of the participants who played the game mostly as a sound game highlighted the importance of visual feedback for seeing the screen and where the fingers were placed. Ten of the participants consider that they could play the game with any single modality (audio or visuals only), with two answering that they would not be able to play without sound.

Regarding usability, participants were asked if anything frustrated them in the game. Nine of the 12 participants mentioned that the strictness of where to tap on the screen to reproduce a given ShapeTones, and the fact that there were no visual aids for this frustrated them. This frustration would increase in higher levels of the game. As one of the participants put it: *"I got the relationship [between the ShapeTones] right but the position wrong – there was a mismatch between my head and the screen"*. Another participant stated that this frustration *"is part of the fun"*. The same frustration was also conveyed when participants were asked to suggest further improvements – six of the participants suggested showing the ShapeTones sections (permanently or only temporarily). We observed that participants used different strategies for solving this issue and achieving a higher precision, by a strict positioning of the hand or by moving the device. Some remarks in the interviews confirm this, e.g. *"I tried to hold it in a different way, shifted and treated it as a piano"*. When asked if they would play the game again, eight of the 12 participants answered affirmatively, with two answering *"maybe"* and two negatively, one of which stating *"I'm not much of a player"* and the other mentioning lack of *"entertainment value"*. Four of the participants mentioned that they would recommend it as a pedagogical game for musical training.

Discussion

The study confirmed our hypothesis S2H1 that the congruent spikes mapping is more engaging than the incongruent arbitrary shapes mapping. The results from the UESz questionnaire point in this direction in all four subscales, although the results from the interviews reveal some slight variations. Mostly, the interviews confirmed the results from the questionnaires, manifesting preference in terms of aesthetics and crossmodal correspondence with spikes and tones. This is illustrated by statements as *"Spikes is more gratifying, easier to play"*. However, one of the participants showed a preference for arbitrary shapes in general, as shapes with a spikes mapping *"were more similar"*. Another participant stated a preference for arbitrary shapes in higher levels of the game (with higher number of ShapeTones) – he argued that in higher sections spikes were harder to disambiguate (it was harder to distinguish shapes with 6 or 7 spikes, for example), while the distinctiveness of arbitrary shapes became more useful. This might point to a problem with recalling the spikes mapping beyond a certain number of spikes. One of the participants who reported preference for the spikes mapping mentioned that the challenge posed by arbitrary shapes could make it more interesting for repeated play. In relation to the perceived importance of audio-visual display in the game, most of the participants (11 of 12) reported relying on both modalities to play the game. Independently of the main modality (audio or visual), most of them (eight) would rely more on the secondary modality as the difficulty increased, as a backup or additional check. This is in line with literature on the importance of audio for user experience in games [20].

Some participants reported frustrations during the study. A common element of frustration was the inability to see the sections, which caused more missed tones as the levels increased. However, one of the participants mentioned that this frustration was *"part of the fun"*. Although the game has a vertical orientation, two of the participants tilted the device, diagonally or horizontally, to better align their hands and fingers with the tablet. When asked about these strategies, one participant mentioned that he was trying to keep a constant hand alignment to the tablet. He observed that accidentally moving the tablet would misalign his hand, leading to a need to *"recalibrate"* his hand. Another participant mentioned that he was trying to align the tablet horizontally as a piano, a musical metaphor which he was familiar with. Two of the participants would hum back in tone a sequence after it was played, and before tapping. When asked why they did this, they replied that it would help memorisation and repetition. It represents a kind of auditory sketching before committing to a sequence. These elements – importance of keeping or removing frustrating elements, spatial strategies outside the frame of the tablet, auditory sketching before playing – could point towards future research directions.

GENERAL DISCUSSION & CONCLUSIONS

In this paper we examined crossmodal congruence in the context of a memory task in which we evaluated users' ability to determine spatial orderings of a sequence of items on the basis of audio, visual and audio-visual stimuli. Two studies were reported which explored task performance and user experience of crossmodal interaction with congruent, incongruent, and semi-congruent displays. In this section we summarise and compare the insights gained from these studies.

Congruent mappings are preferred, but the addition of audio cancelled its advantages: Findings from Study 1 showed that while a congruent spikes mapping led to better results in terms of task performance, its advantages were cancelled out by the addition of audio output. Findings from Study 2, on the other hand, showed that the combination of audio output with a spikes mapping led to more user engagements as measured by UESz. Both studies also revealed problems with the spikes mapping when the complexity of the task increased (levels three, four, and five) and some preferences for the incongruent arbitrary shapes mapping with respect to the challenge and engagement of crossmodal gameplay. Therefore, there could be a threshold at which the clarity and effectiveness of the congruent mapping is saturated. Whilst requiring significantly more time to complete, the distinctive visual characteristics of the shapes used in the arbitrary shapes could provide a better mapping in those cases. Interestingly, the use of the size mapping as a semi-congruent display yielded poor results, even though it was based on crossmodal correspondences between vertical location, pitch and object size [2, 26, 5]. The type of task, in this case recalling the order of a sequence of items, as opposed to identifying a single item, challenged the effectiveness of these particular crossmodal correspondences and therefore calls for more careful design when using this mapping in crossmodal interfaces.

Preference for crossmodal display expressed, but not always confirmed by data: Most of the participants from Study 1 expressed a preference for audio-visual display. In Study 2, the majority of the participants also preferred using both modalities for playing the crossmodal game. However, scores were higher in Study 1 in the visual-only conditions, which contradicted the subjective feedback and observed interaction strategies in both studies. Studies of crossmodal support for spatial attention and motor learning often point out the positive effects of concurrent feedback. However, in general, little work has examined retention tests without audiovisual feedback [32] as we report. The presented studies therefore contribute a systematic evaluation of crossmodal feedback in the context of multimodal information processing. Indeed, our results point toward a subjective preference for crossmodal as opposed to unimodal interaction when task complexity increases. This is evidenced by the diminished effects of levels of congruency observed when auditory output was introduced in the audio-visual displays in Study 1. These findings are inline with accounts of self-management of working memory resources that is associated with multimodal interaction when there is an increase in cognitive demands [25].

Emergence of complimentary strategies using a primary and secondary modality: The above insight is related to a further observation that was also common to the two studies. In both studies, we have seen some users who prefer visuals and others who prefer audio as the primary mode, though both make more use of the secondary mode as task complexity increases. Further research should examine correlations between preferred primary mode and users background and demographics, e.g. musical training or preferred learning style.

Incongruent crossmodal mappings can sometimes be appealing: In both studies, incongruent crossmodal mappings were sometimes associated with positive effects, namely by presenting a challenge and a level of difficulty that rendered the interaction more interesting for some participants. It was *"part of the fun"*. These observations point towards an alternative dimension of crossmodal interfaces when seen from the perspective user experience and engagement, and not merely task performance. Further studies of user engagement through crossmodal interaction should therefore consider addressing this dimension in design.

Contributions, limitations & further research

The presented studies confirmed findings of previous research on the positive performance effect of congruent display for a new task - the memory task. We also found that the addition of auditory display impacts the effects of the levels of congruency and that participants increasingly relied on multiple modalities as task complexity increased. We showed how task and user experience and engagement measures could be used to inform the design of crossmodal interaction which had not been attempted previously. We also demonstrated the deployment of the UESz in a new domain (crossmodal games) where we found it to be an effective measure of engagement.

There are limitations to these findings, however. First, the relatively small number of participants and the specific type of task used in both studies make it unclear how these findings would generalise to other types of interactions. Second, while participants showed superior performances when using the congruent spikes mapping, it is difficult to predict how successful this particular mapping would be for higher levels of complexity, for example when spikes discernibility and hence the ability to count them becomes more challenging as they represent more complex levels (e.g. beyond 10 spikes). Third, while the addition of audio output was perceived as useful, we only used one type of auditory display and did not vary its congruency mappings. It therefore remains unclear how different levels of congruency of the audio output will change the obtained results, for example by using different timbres, or multiple tones that could also be counted to correspond to different levels on the screen. Fourth, we have displayed the ShapeTones such that they are shown in a neutral position on the screen. It would be interesting to examine how displaying ShapeTones in their corresponding sections on the screen would impact participants performances on retention tasks. Finally, in relation to measuring engagement, we have used only one type of questionnaire. Using additional types of measurements could therefore lead to more insights into users engagement with crossmodal displays.

Nonetheless, our findings raise several questions which we would like to explore further. Firstly, further investigation is needed into the relationships between congruity of display, preferred modality, task complexity, and performance. Secondly, explorations of how 'challenging' aspects of crossmodal mappings can be used to enhance playful user experiences are needed. Thirdly, exploring how the role of crossmodal elements outside the device, such as proprioceptive mappings, could inform the design of engaging crossmodal interaction. Finally, our long term aim is to explore how crossmodality could be used to inform the design of engaging experiences for people with a variety of sensory capabilities.

ACKNOWLEDGMENTS

This work was supported by EPSRC grant EP/J017205/1 and EU Marie Curie fellowship FP7 REA grant 627922.

REFERENCES

1. P Bach-y Rita. 1988. Brain plasticity. *Rehabilitation Medicine. Mosby* (1988), 113–8.
2. Elisheva Ben-Artzi and Lawrence E Marks. 1995. Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics* 57, 8 (1995), 1151–1162.
3. Elizabeth A. Boyle, Thomas M. Connolly, Thomas Hainey, and James M. Boyle. 2012. Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior* 28, 3 (May 2012), 771–780. DOI: http://dx.doi.org/10.1016/j.chb.2011.11.020
4. Alfred O Effenberg. 2005. Movement sonification: Effects on perception and action. *IEEE multimedia* 2 (2005), 53–59.
5. Karla K Evans and Anne Treisman. 2010. Natural cross-modal mappings between visual and auditory features. *Journal of vision* 10, 1 (2010), 6.
6. Thomas Ferris, Robert Penfold, Shameem Hameed, and Nadine Sarter. 2006. The implications of crossmodal links in attention for the design of multimodal interfaces: A driving simulation study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. Sage Publications, 406–409.
7. Mark A Guadagnoli and Timothy D Lee. 2004. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of motor behavior* 36, 2 (2004), 212–224.
8. Stephen J Guastello, Katherine Reiter, Matthew Malon, and Anton Shircel. 2015. When auditory and visual signal processing conflict: cross-modal interference in extended work periods. *Theoretical Issues in Ergonomics Science* 16, 3 (2015), 232–254.
9. Eve Hoggan, Topi Kaaresoja, Pauli Laitinen, and Stephen Brewster. 2008. Crossmodal congruence: the look, feel and sound of touchscreen widgets. In *Proceedings of the 10th international conference on Multimodal interfaces*. ACM, 157–164.
10. Michael Xuelin Huang, Will WW Tang, Kenneth WK Lo, CK Lau, Grace Ngai, and Stephen Chan. 2012. MelodicBrush: a novel system for cross-modal digital art creation linking calligraphy and music. In *Proceedings of the Designing Interactive Systems Conference*. ACM, 418–427.
11. Joseph 'Jofish' Kaye, Kirsten Boehner, Jarmo Laaksolahti, and Anna Stahl. 2007. Evaluating experience-focused HCI. In *CHI '07 extended abstracts on Human factors in computing systems (CHI EA '07)*. ACM, New York, NY, USA, 2117–2120. DOI: http://dx.doi.org/10.1145/1240866.1240962
12. Ju-Hwan Lee and Charles Spence. 2008. Feeling what you hear: Task-irrelevant sounds modulate tactile perception delivered via a touch screen. *Journal on Multimodal User Interfaces* 2, 3-4 (2008), 145–156.
13. Dominic W Massaro. 1998. Illusions and issues in bimodal speech perception. In *AVSP'98 International Conference on Auditory-Visual Speech Processing*.
14. Harry Mcgurk and John Macdonald. 1976. Hearing lips and seeing voices. *Nature* 264, 5588 (Dec. 1976), 746–748. DOI: http://dx.doi.org/10.1038/264746a0
15. Peter BL Meijer. 1992. An experimental system for auditory image representations. *Biomedical Engineering, IEEE Transactions on* 39, 2 (1992), 112–121.
16. Oussama Metatla, Nick Bryan-Kinns, Tony Stockman, and Fiore Martin. 2012. Supporting cross-modal collaboration in the workplace. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers*. British Computer Society, 109–118.
17. Oussama Metatla, Nick Bryan-Kinns, Tony Stockman, and Fiore Martin. 2015. Sonification of reference markers for auditory graphs: Effects on non-visual point estimation tasks. In *PeerJ PrePrints*.
18. Micah M Murray, Sophie Molholm, Christoph M Michel, Dirk J Heslenfeld, Walter Ritter, Daniel C Javitt, Charles E Schroeder, and John J Foxe. 2005. Grabbing your ear: rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex* 15, 7 (2005), 963–974.
19. Shaun Musgrave. 2015. 'Dark Echo' Review - Silence Is Golden, And So Is This Game. (Feb. 2015). http://toucharcade.com/2015/02/11/dark-echo-review/
20. Lennart E. Nacke, Mark N. Grimshaw, and Craig A. Lindley. 2010. More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers* 22, 5 (Sept. 2010), 336–343. DOI: http://dx.doi.org/10.1016/j.intcom.2010.04.005
21. Kent L. Norman. 2013. GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers. *Interacting with Computers* 25, 4 (July 2013), 278–283. DOI: http://dx.doi.org/10.1093/iwc/iwt009
22. Heather O'Brien and Paul Cairns. 2015. An empirical evaluation of the User Engagement Scale (UES) in online news environments. *Information Processing & Management* 51, 4 (July 2015), 413–427. DOI: http://dx.doi.org/10.1016/j.ipm.2015.03.003
23. Heather L. O'Brien and Elaine G. Toms. 2008. What is User Engagement? A Conceptual Framework for Defining User Engagement with Technology. *J. Am. Soc. Inf. Sci. Technol.* 59, 6 (April 2008), 938–955. DOI: http://dx.doi.org/10.1002/asi.v59:6

24. Heather L. O'Brien and Elaine G. Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (Jan. 2010), 50–69. DOI:http://dx.doi.org/10.1002/asi.21229

25. Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford. 2004. When do we interact multimodally?: cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 129–136.

26. Geoffrey R Patching and Philip T Quinlan. 2002. Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance* 28, 4 (2002), 755.

27. Mary C. Potter, Brad Wyble, Carl Erick Hagmann, and Emily S. McCourt. 2013. Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics* 76, 2 (Dec. 2013), 270–279. DOI: http://dx.doi.org/10.3758/s13414-013-0605-z

28. Rick Raymer. 2011. Gamification: Using Game Mechanics to Enhance eLearning. *eLearn* 2011, 9 (Sept. 2011). DOI: http://dx.doi.org/10.1145/2025356.2031772

29. Paul Rodway. 2005. The modality shift effect and the effectiveness of warning signals in different modalities. *Acta Psychologica* 120, 2 (2005), 199–226.

30. Richard A Schmidt. 1991. Frequent augmented feedback can degrade learning: Evidence and interpretations. In *Tutorials in motor neuroscience*. Springer, 59–75.

31. Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. 2000. Illusions: What you see is what you hear. *Nature* 408, 6814 (Dec. 2000), 788–788. DOI: http://dx.doi.org/10.1038/35048669

32. Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic bulletin & review* 20, 1 (2013), 21–53.

33. Charles Spence. 2010. Crossmodal spatial attention. *Annals of the New York Academy of Sciences* 1191, 1 (2010), 182–200.

34. Charles Spence. 2011. Crossmodal correspondences: A tutorial review. *Attention and Perceptual Psychophysics* (2011). DOI: http://dx.doi.org/DOI10.3758/s13414-010-0073-7

35. Charles Spence and Jon Driver. 1997. Cross-modal links in attention between audition, vision, and touch: Implications for interface design. *International Journal of Cognitive Ergonomics* (1997).

36. Charles Spence and Jon Driver. 2004. *Crossmodal space and crossmodal attention*. Oxford University Press.

37. Charles Spence and Cristy Ho. 2015. Multisensory information processing. In *APA handbook of human systems integration*, D. A. Boehm-Davis, F. T. Durso, and J. D. Lee (Eds.). American Psychological Association, Washington, DC, Ch. 27.

38. JK Stefanucci and DR Proffitt. 2005. Multimodal interfaces improve memory. In *Proceedings of the 11th International Conference on Human-Computer Interaction (HCI'05)*.

39. Penelope Sweetser and Peta Wyeth. 2005. GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)* 3 (July 2005), 3–3. DOI:http://dx.doi.org/10.1145/1077246.1077253 ACM ID: 1077253.

40. Andrew Webster. 2013. Gaming in darkness: 'Papa Sangre II' is a terrifying world made entirely of sound. (Oct. 2013). http://www.theverge.com/2013/10/31/5048298/papa-sangre-ii-is-a-terrifying-world-made-of-sound

41. Eric N. Wiebe, Allison Lamb, Megan Hardy, and David Sharek. 2014. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior* 32 (March 2014), 123–132. DOI: http://dx.doi.org/10.1016/j.chb.2013.12.001

42. Fredrik Winberg. 2006. Supporting cross-modal collaboration: adding a social dimension to accessibility. In *Haptic and Audio Interaction Design*. Springer, 102–110.

43. Peter Wright, Jayne Wallace, and John McCarthy. 2008. Aesthetics and experience-centered design. *ACM Transactions on Computer-Human Interaction* 15, 4 (Dec. 2008), 18:1–18:21. DOI: http://dx.doi.org/10.1145/1460355.1460360

USER-CENTERED DESIGN OF A TOOL FOR INTERACTIVE COMPUTER-GENERATED AUDIOVISUALS

NUNO N. CORREIA
Goldsmiths, University of London
London, United Kingdom
n.correia@gold.ac.uk

ATAU TANAKA
Goldsmiths, University of London
London, United Kingdom
a.tanaka@gold.ac.uk

ABSTRACT

The present study aims to design a tool for interactive computer-generated audiovisuals. In this paper, we investigate if the tools for audiovisual performance and composition have caught up with the growing interest and the practices in the field. We have adopted a user-centered design approach for our study, based on interviews and a workshop with practitioners. The interviews identified key themes – expressivity, ease of use and connection with the audience – that were explored in the workshop. During the workshop, a novel methodology was adopted – *reboot* – which expands upon the *bootlegging* technique. Key ideas regarding audiovisual performance gathered from the interviews; sketches for novel audiovisual tools resulting from the workshop; and the reboot technique, are the main contributions of this study.

KEYWORDS

User-Centered Design, Interaction Design, Human-Computer Interaction, Audiovisual Performance, VJing, Computer-Generated Graphics, Computer-Generated Sound.

1. INTRODUCTION

The field of audiovisual (AV) performance and composition has been particularly active in recent years. New festivals (for example: LPM,[1] LEV,[2] Mapping[3]), publications (for example: *See This Sound* series and web archive,[4] *LEA Live Visuals* special issue[5]) and conferences/seminars (for example: Seeing Sound,[6] Real-Time Visuals[7]), have focused in this field in the last years. From our own experience as performers, we have realized that audiovisual performances often rely on custom software made by the artists, and not on ready-made tools available to other performers. We would like to understand if the tools for AV performance and composition have caught up with the growing interest and practices in the field. The practical aim of this study is to design a tool for computer-generated audiovisuals, taking into account expressiveness, ease of use, and audience involvement. In this context, we consider that expressiveness is "not a distinct action or task that can be isolated for study, but rather a phenomenon that arises as a consequence of how an action is completed" (Hook et al. 2011). In this paper, we present early results from research examining user interfaces for procedural audiovisual performance systems.

We adopted a User-Centered Design (UCD) approach consisting of two steps. We first conducted interviews with 12 audiovisual practitioners, to better understand their practice, in particular: the strengths and weaknesses of the tools that they use; and the role of the audience in their performances. We then conducted a 1-day workshop to brainstorm, create imaginary scenarios, and sketch possible future tools for audiovisual performance, taking into account themes identified in the previous interview stage. 19 participants attended the workshop. During the workshop, we implemented the *bootlegging* brainstorming methodology (Holmquist 2008) and introduced a novel twist on it, which we named *reboot*. This study gave rise to: key ideas on tools for audiovisual performance gathered in the interviews; the sketches for a novel tool for AV performance produced in the workshop (which used the key ideas as an input); and the reboot method (which was devised as a means to rapidly generate sketches based on an initial input).

1. LPM: *http://liveperformersmeeting.net*
2. LEV: *http://www.levfestival.com*
3. Mapping: *http://www.mappingfestival.com*
4. See This Sound: *http://see-this-sound.at*
5. LEA Live Visuals special issue: *http://www.leoalmanac.org/vol19-no3-live-visuals/*
6. Seeing Sound: *http://www.seeingsound.co.uk*
7. Real-Time Visuals: *http://www.realtimevisuals.org*

2. TOOLS FOR INTERACTIVE AUDIOVISUALS

Audiovisual performance has a long history, from color organs and the visual music cinema performances of early 20th century pioneers – artists such as Walther Ruttmann and Oskar Fischinger, who used "tinted animation to live musical accompaniment" (Moritz 1997) – to contemporary digital works. From the 1990s, there has been a strong interest in "screen-based performance", adopting "a long litany of names such as audiovisual performance, real-time video, live cinema, performance cinema, and VJ culture" (Salter 2010, 171). Chris Salter attributes this interest to two branches of techno-cultural development: on the one hand, "breakthroughs in digital computation, particularly the development of hardware and software components for the capture, processing, and manipulation of image and sound" and on the other hand, "the international rise of the techno/club scene, which rapidly exploited such technologies". From the terminology mentioned by Salter, we preferentially use *audiovisual* or *AV performance*, as it best encapsulates the two modalities of sound and graphics.

Two notable examples of contemporary audiovisual artists using computer-generated graphics and sound are Golan Levin and Toshio Iwai. They are relevant to this study because they are concerned with creating interfaces and instruments for audiovisual expression. Levin developed a suite of works under the name *Audiovisual Environment Suite* (*AVES*) and described his approach to audiovisual performance as being based on painterly interfaces (Levin 2000). Iwai creates playful pieces, crossing genres between game, installation, performance (with works such as *Elektroplankton*, *Composition on the Table*) and audiovisual instrument (with *Tenori-On*)(Wynne 2008).

There is a large choice of software tools for audiovisual performance. In this context, we use the term "tool" to define generic software systems that can be used by different artists to create their own performances (and not software created by an artist for a specific piece). These tools deal with audio, visuals or both. They can be ready-made commercial software such as Modul8,[8] Resolume,[9] VDMX[10] (with an emphasis on graphics) or Ableton Live[11] (with an emphasis on sound). There are also open-ended programming frameworks or environments – usually following either data-flow programming or textual programming paradigms. They usually carry with them steeper learning curves than turnkey software products. Examples of data-flow programming software used for audiovisual performance: VVVV,[12] Quartz Compos-

8. Modul8: *http://www.modul8.ch*

9. Resolume: *http://resolume.com*

10. VDMX: *http://vidvox.net*

11. Ableton Live: *https://www.ableton.com*

12. VVVV: *http://vvvv.org*

er[13] (with an emphasis on graphics), PureData[14] (emphasis on sound) and Max/MSP/Jitter.[15] Examples of textual programming frameworks or environments used for audiovisual performances: SuperCollider[16] (mainly for sound), openFrameworks[17] and Processing.[18]

Most ready-made commercial software tools for live visuals (such as Modul8, Resolume and VDMX) focus on video playback and manipulation. Therefore, artists interested in using video for their performances have a choice of using either ready-made (and easier to use) software, or programming languages / environments (with a steeper learning curve, but offering more flexibility). For artists dealing with computer-generated graphics, however, there is a scarcity of ready-made, easy to use software.

The design of tools for AV and VJ (Video Jockey) performances has been analyzed before from these perspectives: taking into account expressive interaction (Hook et al. 2011); ease of use (Correia and Kleimola 2014); and audience, specifically considering participation (Taylor et al. 2009) and awareness of performer's actions (Lew 2004). Our work is distinct because it takes into account all three aspects; it focuses on computer-generated audio and visuals; and because of the novel methodological approach regarding user-centered design.

3. METHODOLOGY

This study follows a UCD approach. UCD is "a broad term to describe design processes in which end-users influence how a design takes shape" (Abras, Maloney-krichmar, and Preece 2004). In this case, the end-users are audiovisual performers. We adopted a UCD approach to better understand current practices of audiovisual performers and to design a tool that addresses their needs. The interviews aimed to obtain insights into the practices of audiovisual performers, and the tools they use. The questions were grouped in six sections:

- Characterization of performer;
- Tools;
- User Interface (UI);
- Audience involvement;
- Artistic goals and technology; and
- Specific performance recollection.

The interviews were conducted prior to the workshop, so that the insights gathered during the interview stage could inform the scenarios for the workshop. Workshops are defined as "collaborative design

13. Quartz Composer: *http://quartzcomposer.com*
14. PureData: *http://puredata.info*
15. Max/MSP/Jitter: *http://cycling74.com*
16. SuperCollider: *http://supercollider.sourceforge.net*
17. openFrameworks: *http://www.openframeworks.cc*
18. Processing: *https://processing.org*

events providing a participatory and equal arena for sharing perspectives, forming visions and creating new solutions” (Soini and Pirinen 2005). Due to the collaborative and participatory nature of workshops, they were chosen as a key element of the adopted methodology. A one-day, 6-hour workshop was conducted, aiming to produce sketches of novel tools for audiovisual performance.

For the first part of the workshop, we conducted a *bootlegging* session. Bootlegging is a “structured brainstorming technique particularly suited to multidisciplinary settings” (Holmquist 2008, 158). Bootlegging applies the notion of cut-up – a form of literary collage popularized by William Burroughs – to brainstorming sessions, mixing familiar concepts in a way that stimulates creativity. A bootlegging session requires a theme. It also requires the definition of four categories for idea generation, two relative to the user side and two related to the theme and technology. A presentation format should also be chosen. The participants, divided into groups, should then generate several ideas (as post-its) for each category, mix those ideas and create 4-5 random combinations of each category per group. Those combinations then become the trigger of a brainstorming session, attempting to imagine different potential applications for each combination. Afterwards, the groups are asked to pick one of the ideas and prepare a presentation in the chosen format (Holmquist 2008, 159).

For the second half of the workshop, we devised and ran a variation of the bootlegging technique, which we entitled *reboot*. Reboot is a brainstorming technique that builds upon bootlegging, and is intended as a follow-up to a bootlegging session. Similarly to bootlegging, it also requires a theme and four categories (the same ones as in the preceding bootlegging session) for idea generation. For more focused results, additional requirements are introduced to the initial theme, taking into account the results of the bootlegging session. Instead of relying on generating multiple variables for each category and random mixing, the variables for the four categories are deliberately chosen by the participants (one variable per category). Some or all of these variables may also be defined by the session facilitators. The same steps as in bootlegging are taken, with the exception of the mixing and combining steps. The aim of reboot is to give direction and focus after the open-ended and aleatoric nature of the first exercise. After having stimulated the creation of new application ideas with the bootlegging session, reboot allows the participants to concentrate on more specific solutions.

4. INTERVIEWS

4.1. PARTICIPANTS

We conducted 12 face-to-face interviews lasting between 25 and 56 minutes. 11 of the interviewees were male, 1 was female. The interviewees had between 4 and 18 years of performance experience.

4.2. RESULTS

When asked what is the most important feature of the tools they use, two interviewees mentioned modularity and flexibility of the software (“easily adaptable to different performance situations and its flexibility”; “the fact that it can be configured in so many different ways”). Two artists mentioned ease of integration with hardware and other software (“the way that Modul8 is built, with the options that you have, basically controlling those options with knobs and faders” and “Resolume was always working well alongside Ableton”). Two others mentioned expressivity and fluidity (“it creates images a bit more like you were creating music”; “you want to be like a musician, you want to play an instrument, you want to respond in real-time”). Other interviewees mentioned integration of environmental elements (“construction with the elements that are around”), generative capabilities and diversity (“the fact that it’s generative (...) each performance becomes different”), communication of live creative process to the audience (“projecting agency to the audience”), reliability (“software can be glitchy, slow, crash”) and speed (“I want to be able to do multiple processes very quickly”).

When asked what features they would like to add to their performance tools, interviewees repeated qualities mentioned earlier, such as stability, modularity and diversity. Additionally, two artists mentioned that they would like to have a flexible timeline view in their software (because “the time of the performance is of a different time from the reality” and “for running more generative kind of installation type stuff”). Ease of mapping audio reactivity to graphics was also mentioned (“the ability to make a video file or a layer audio reactive with a single button”).

Regarding ease of use, the interviewees who use commercial software agreed that these tools are easy to use. The others consider that the custom systems they have built are personal and not designed for others to use (“we always get it quite personal”; “I don’t care about ease of use I care about expressiveness”; “I don’t think that the system itself is complicated but the way it’s controlled might be complicated”; “it’s more the realization that it is your own tool and that you’re showing your composition through that tool where the value lies”). Two of the artists make a distinction between systems created for their own performances, focusing on expressiveness and individuality, and systems that they have created for others, which are easier to use.

Regarding preference for type of UI, nine of the 12 interviewees use hardware controllers (with two expressing a preference for motorized controllers), and five of these complement the hardware controller with an Apple iPad running a controller software application (app). Hardware controllers and iPad (running Touch OSC or Lemur apps) are used to control the audio and/or visual software running on the laptop. Hardware controllers are favored because of the eyes-off tactile feedback they provide. The following quote reflects a general view for a majority of the interviewees: “the physical feeling for me is essential for performance: buttons, rotaries whatever; because I’m more precise – they never let me down and I feel the performance better”. For some, motorized controllers are preferred: “a motorized physical controller with real sliders makes it easier to be able to look at the screen without the need of looking at the controller”. iPads are used because of the identification and visual information they provide: “it’s really an easy way of labeling up all your effects and be able to see all that stuff without having to stick all bits of plastic to MIDI controllers or to keys in your keyboard”, although that comes with a cost: “but of course the problem is that you need to be looking at the iPad because you don’t feel with the finger”.

One of the artists uses live coding as a performance technique, because in his opinion “graphical interfaces are frustrating” and slow. He considers live coding natural for him, as he uses SuperCollider. He has some doubts regarding the impact of live coding on the audience: “I have a bit of a problem with live coding and people showing the screen, you know – I always just stand there and wonder how it’s like for most people”. The solution he has found is to integrate the code with the visuals: “I’m trying to find creative ways to display the code and also make it part of the graphics”. Another interviewee explores showing the Graphical User Interface (GUI) as a means of projecting the performance process to the audience: “there’s two visuals going on, there’s the visual object that is showing, which is somehow the thing to be manipulated, and then there’s the act of manipulation itself, which is some kind of GUI that sits on top of that”. He tries to find a balance between having more GUI and more ease of use for him, or less GUI and therefore less visual interference for the audience: “I could put loads of GUI and make things maybe clearer for the audience and they could see more of my actions, but then it starts to crowd over the graphics that are underneath”. The remaining controls are executed with key presses. Two other artists use only the computer keyboard and keyboard shortcuts as their interface.

4.3. AUDIENCE REACTION AND PERCEPTION OF LIVENESS

Audience reaction to the performance, as perceived during the performance or communicated afterwards, is important for eight of the 12 interviewees. When questioned if their audiences understand the interactive and real-time element of the performances, five replied that it depends on the audience and the setting. According to these artists, some audiences might be more knowledgeable in computer-based performance than others, whereas in some venues the visual element might not be as valued as in others. Four of the artists state that it is indifferent for them if the audience understands that the visuals are interactive or not. For these artists, the importance of the performance lies in the quality of the experience, not in the perception that it is live. For two of the interviewees, audience perception of liveness derives from the assumption that it is live if there is someone on stage ("if you see … another people doing other things") or to post-performance feedback ("they'll actively tell me why they've enjoyed it … I'm pretty confident that it's communicating what it's trying to"). One interviewee considers that the audience generally does not understand that the performance is being done live – "people can't see much what we're doing" and "people think once you have a laptop on stage that laptop is doing everything for you", therefore: "we are considering: should we actually make that clearer".

Interviewees were asked to suggest ways to improve audience understanding of liveness. Two of the interviewees did not have interest in improving communication with the audience, with an additional one stating that it would make sense only in specific performances. Live coding, or further displaying aspects of the code, is a possible path for four of the artists. The live coding interviewee suggests further integration between displaying code and additional visuals ("make the codes animated somehow" and "add some comedy to it"). Two artists who are not currently using live coding contemplate using that performance technique in future work. Another interviewee mentioned the notion of "debug interface" to showcase parameters to the audience, in the same way that an artists uses debug windows to check for values ("almost like another layer of visual information that's purely only really for the developer but that is displayed for the audience"). Two of the artists suggest adding live camera feeds to convey a sense of liveness, either pointed to the audience ("more cameras where the space of the audience is") or to their stage setup ("a camera over my head on my set up showing what I'm doing"). Additional suggestions are: using custom apps that the audience could download and interact using their mobile devices during a performance ("custom apps or information that's being kind of gathered or created by the audience); and tracking audience movement as an interaction mechanism ("body positioning, and somehow one of the persons in the audience can affect the music somehow, or the visuals").

5. WORKSHOP

5.1. PARTICIPANT CHARACTERIZATION

The one-day workshop took place in October 2014, at Goldsmiths, University of London. The call for participation was circulated among mailing lists within the Goldsmiths and London Video Hackspace[19] communities. 19 participants (12 male and 7 female) took part in the workshop. Ten described themselves as VJs and/or AV performers, three as programmers, one as video artist, and four as musicians - all practitioners in the field of audiovisual performance or related fields (music, video, media arts). One anthropologist studying audiovisual performance also participated in the workshop. Four of the partici-

19. London Video Hackspace: *http://www.videohackspace.com*

pants develop work with video footage, another four with computer-generated graphics and six with both. Nine of the participants stated that they build their own tools for performance, with Max/MSP (five), openFrameworks (three) and with Processing (one). Three of the workshop participants had been interviewed in the previous stage of the study.

5.2. BOOTLEGGING

In our bootlegging session, the theme was: "Software for interactive computer-generated audiovisuals, using a single screen". The constraint of the single screen aimed to stimulate creativity in terms of user interface, avoiding a performer-specific screen populated with GUI, common in commercial software. The participants were divided into five groups. During the generation stage, each group produced post-its with dozens of variables for each of the chosen categories – user, situation, interface and device. In the mixing stage, these were randomly mixed within each group, and each group was asked to produce four random combinations with one item per category. Each of these combinations was pasted to an A3 paper. The groups were then asked to think of different applications per combination. Finally, they were asked to pick one of the applications and develop it conceptually, preparing a presentation based on a storyboard and wireframes (figure 1).

The bootlegging session achieved the aim of stimulating creativity in participants and opening up the range of possibilities for audiovisual performance outside of the usual scenarios. Many of the concepts were humorous, ironic and playful. The five concepts were:

- *Botanical garden motion sensors*, a garden transformed into a performance space, augmented with surround sound and visuals projection-mapped on trees;
- *Fish food - an audio-fishual dance ensemble*, a reactive aquatic audiovisual environment for public spaces;
- *Interactive surgery blanket*, a special fabric for health purposes, incorporating a flexible screen, which reveals physiologic aspects of the patient it is covering, with bodily functions being sonified and visualized;
- *EAVI sleeper*, a system incorporating a blanket with different biological sensors, which generates an audiovisual performance based on the biological data of a sleeping "performer"; and
- *Blind date sensory experience*, a system for two artists who meet on an online "blind date" for a networked audiovisual performance.

Figure 1 Bootlegging presentation

5.3. REBOOT

After the serendipity, humor and technological speculation generated by the bootlegging stage, the reboot stage aimed to bring more focused results. The participants were regrouped into different combinations. The groups were asked to brainstorm on the same theme as the bootlegging session, but adding a few more constraints:

- to focus on a performance scenario, and
- to take into account key qualities in tools for audiovisual performance detected during the interviews – expressivity; ease of use; and connection with the audience.

After the brainstorming session, the groups were asked to prepare a presentation, also based on a storyboard and wireframes.[20] Two of the concepts (*Gestural Touchscreen* and *Meta/Vis*) aimed to reach a balance between expressivity and ease of use. The additional three concepts focused on audience participation. Two of these (*Sensor Disco* and *Fields of Interference*) consist of performance spaces without a single main performer – the audience becomes the performer:

- *Gestural Touchscreen* is a touch-screen based application, controlled entirely by gestures. There is no GUI. Users can only load SVG files as visual content and there is a built-in physics engine (figure 2).
- *Meta/Vis* also relies on multitouch, but adds a "pre-performance" configuration stage. This stage adopts a data-flow paradigm, although substantially simplified. Objects such as sound, visuals, control, generative and physics can be linked with arrows in different configurations, and contain drop-down menus for additional options. The group described it as "a simplified Jitter-style patching system".

20. The five sketches can be downloaded from: *http://nunocorreia.com/files/IG-AV-sketches.zip*

- *Sensor Disco* consists of an environment containing multiple sensors. By moving in the space, audience members trigger and modulate sounds, which are visualized on the walls and on the floor.
- In *Fields of Interference* users creates sound and visuals by moving with their mobile devices in a room. The system is composed of an array of sensors, which sonifies and visualizes Wi-Fi interference from mobile devices – using surround sound and an immersive dome-like projection screen.
- In *Beat the DJ*, there is a main performer role (in this case, a DJ/VJ), and the club environment becomes a game where audience activity "unlocks" audiovisual content. In the beginning, the audio and visuals are simple (for example, a drum loop and a few melody lines) but audience reaction can give the DJ/VJ more elements to play with. These elements can potentially trigger further reactions from the audience.

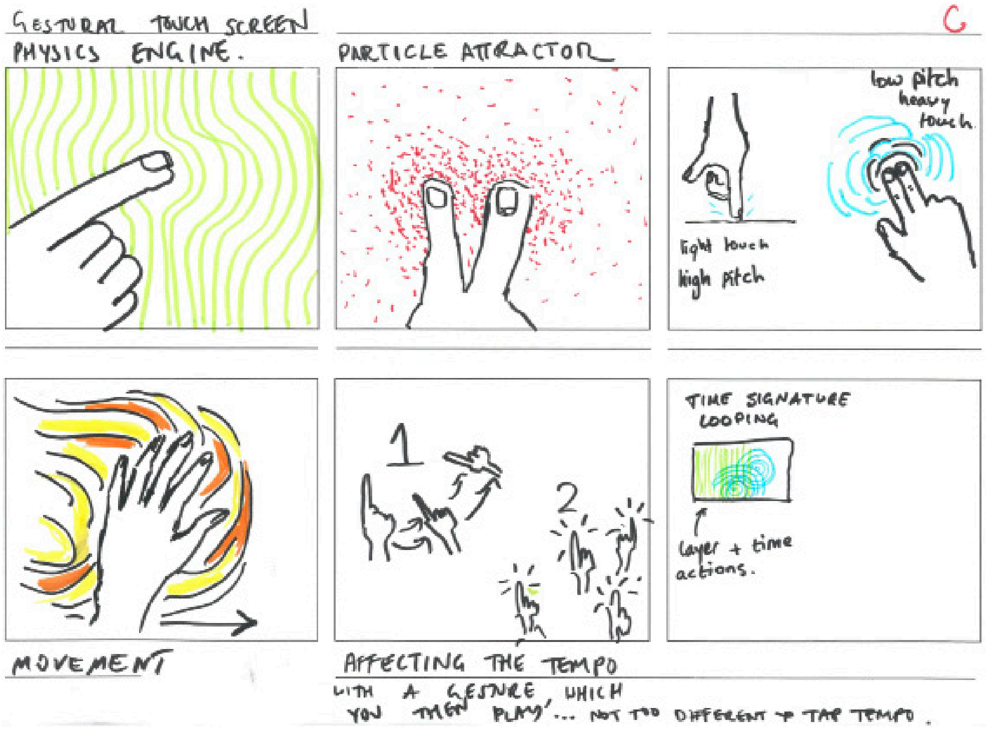


Figure 2 Storyboard from reboot session (*Gestural Touchscreen*)

6. DISCUSSION

The adoption of a UCD approach generated surprising results, which would not have been achieved from a top-down design process. In the beginning of the reboot session, we asked participants to reflect upon themes identified in the interview stage – expressivity, ease of use and connection with the audience. The resulting sketches successfully incorporated those reflections. The unconventional approaches of several of the sketches would not have been possible without the earlier bootlegging session, which stimulated out of the box thinking amongst the participants, enabling them to envision possibilities that go beyond traditional solutions. We were thus satisfied with the method taken, from interview and identification of themes to bootlegging and reboot. We believe that reboot is an important methodological contribution of the study.

6.1. EXPRESSIVITY, FLEXIBILITY AND EASE OF USE

One of the key themes detected in the interviews was expressivity, to be able to make visuals "like a musician" and the desire to play an audiovisual tool with the same expressivity and fluency as a traditional musical instrument. Another was flexibility and the possibility of reconfiguring the software in many ways. Yet another was ease of use – existing ready-made tools are easy to use, but they focus mostly on video manipulation, and there are few targeting computer-generated graphics. Combining these elements can be challenging, and often there are trade-offs between expressivity, flexibility and ease of use. Two of the sketches that came out of the workshop, *Meta/Vis* and *Gestural Touchscreen*, address these issues. Both rely on multitouch interaction so as to convey a sense of immediate control of sound and visuals. In *Gestural Touchscreen*, the expressivity comes from the rich variety of gestures that can be used to control sound and visuals and from the pressure sensitivity capabilities. The flexibility arises from the possibility of loading SVG (Scalable Vector Graphics) files as visual patterns to be animated and manipulated, making the graphical possibilities virtually endless. *Meta/Vis* also relies on multitouch gestures for expressivity (although less than *Gestural Touchscreen*). The focus of *Meta/Vis* is on flexibility and reconfiguration. To solve this, while maintaining ease of use, it incorporates a simplified data-flow programming component – basic blocs such as sound, visuals and control that can be re-routed and that contain simple drop-down menus with options. Both *Meta/Vis* and *Gestural Touchscreen* address ease of use by: implementing multitouch gestures that are easy to understand, while allowing for a great variety of control (particularly in *Gestural Touchscreen*); and adopting ingenuously easy solutions for reconfiguration (with the SVG approach in *Gestural Touchscreen*, and the simple data-flow modules of *Meta/Vis*).

6.2. AUDIENCE INVOLVEMENT

Another key theme detected in the interviews was audience involvement: the importance for some artists of conveying the liveness of the performance to audiences; and how to have audiences participate in the performance. Three of the sketches from the workshop address the issue of audience participation. In *Sensor Disco*, audience positioning in the space affects sound and visuals; in *Beat the DJ* the amount of physical activity of audience participation enriches the sound and visuals with a game-like "levels" logic; and in *Fields of Interference* the Wi-Fi signal from mobile phones of audience members is sonified and visualized.

7. CONCLUSIONS

Although the field of audiovisual performance has a long history, it has not been thoroughly documented, and it has not been the subject of design research. Technological developments present numerous opportunities – in interaction with the tools; creation of sound and graphics; visual and auditory diffusion; use of networks; ubiquitous computing; and audience participation. This study focused on one aspect of content generation – computer-generated audiovisuals – and arrives to concepts that explore some of these opportunities for performance, using a UCD approach. The study is an early stage part of our research. With this study, we were able to identify key ideas on audiovisual performance in the interviews; participants produced sketches for novel tools in the workshop; and we conceived and tested the reboot brainstorming technique. The sketches produced in the workshop show great promise in addressing key themes and concerns identified during interviews to practitioners – such as expressivity, flexibility, ease of use and audience involvement. These concepts can be useful for audiovisual performers, or designers of tools for audiovisual performance. The study also proposes an extension to the bootlegging methodology, which we entitled reboot. Reboot extends open-ended brainstorming to bring additional focus to brainstorm sessions through focused iteration. In this case, the focus was defined based on key themes identified during the earlier interviews stage. The interviews set themes. Bootlegging facilitates serendipity and out of the box thinking. Reboot brings themes from interviews into an iteration of bootlegging to provide focus and structure to the brainstorming process without constraining it to a task-based exercise.

In a future stage of the research, we will conduct another workshop with performers and programmers, in order to develop these sketches into functioning prototypes. Some features from the different concepts might be merged into one or more prototypes. Afterwards, we will conduct tests with these prototypes in a performance setting. The prototypes will be made available as open-source code. With this study, we hope to contribute to the audiovisual performance community, and the expansion of the range of creative possibilities at their disposal.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n° 627922. We would like to thank Alessandro Altavilla for assistance with the photo and video documentation.

REFERENCES

Abras, Chadia, Diane Maloney-krichmar, and Jenny Preece. 2004. "User-Centered Design." In I*n Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*. Publications.

Correia, Nuno N., and Jari Kleimola. 2014. "Web Browser as Platform for Audiovisual Performances." In *Proceedings of the 11th International Conference on Advances in Computer Entertainment Technology (ACE)*. Funchal: ACM Press.

Holmquist, Lars Erik. 2008. "Bootlegging: Multidisciplinary Brainstorming with Cut-Ups." In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008*, 158–61. PDC '08. Indianapolis, IN: Indiana University.

Hook, Jonathan, David Green, John McCarthy, Stuart Taylor, Peter Wright, and Patrick Olivier. 2011. "A VJ Centered Exploration of Expressive Interaction." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1265–74. CHI '11. New York, NY: ACM.

Levin, Golan. 2000. *Painterly Interfaces for Audiovisual Performance*. Cambridge, MA: M.Sc. Dissertation. Massachusetts Institute of Technology. *http://acg.media.mit.edu/people/golan/thesis/thesis300.pdf*.

Lew, Michael. 2004. "Live Cinema: Designing an Instrument for Cinema Editing as a Live Performance." In *Proc. NIME 2004*. NIME. http://www.nime.org/2004/NIME04/paper/NIME04_3A03.pdf.

Moritz, William. 1997. "The Dream of Color Music, And Machines That Made It Possible." *Animation World Magazine*, April. *http://www.awn.com/mag/issue2.1/articles/moritz2.1.html*.

Salter, Chris. 2010. *Entangled: Technology and the Transformation of Performance*. Cambridge, MA: MIT Press.

Soini, Katja, and Antti Pirinen. 2005. "Workshops - Collaborative Arena for Generative Research." In *Proceedings of the Conference Designing Pleasurable Products and Interfaces*. Eindhoven.

Taylor, Stuart, Shahram Izadi, David Kirk, Richard Harper, and Armando Garcia-Mendoza. 2009. "Turning the Tables: An Interactive Surface for Vjing." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1251–54. CHI '09. New York, NY: ACM.

Wynne, Graham. 2008. *Analyse the Extent to Which the Electroplankton and Tenori-on Systems Overcome Barriers to Creativity in Modern Music Production*. Independent dissertation. January 2008. Essay. *http://tenori-on.yamaha-europe.com/uk/downloads/GrahamWynneDissertation.pdf*.

AV Zones – Tablet App for Audiovisual Performance

Nuno N. Correia

Goldsmiths, University of London, London, United Kindgdom
n.correia@gold.ac.uk

Abstract. We have identified potential for tablets to be used as stand-alone tools for audiovisual performance, and not simply as controllers, due to their portability, expressive capabilities of multi-touch, and processing power. To explore this potential, we have developed AV Zones (AudioVisual Zones), an iPad app for audiovisual performance. In a preliminary phase, we conducted interviews with audiovisual performers and a workshop, to understand user needs and desires. We then developed AV Zones, an iPad app for audiovisual performance, composed of an audio sequencer/looper with a visualizer. It explores the interactive potential of a touch screen tablet for integrated musical and visual expression. By default, 3 audiovisual columns or "zones" allow for the manipulation of 3 audio loops. These zones are metaphorical adaptations of channels in a standard audio mixer.

Keywords: audiovisual, multi-touch, tablet, multisensorial, touchscreen.

Introduction

The popularity of multi-touch tablets, particularly since the introduction of the Apple iPad in 2010, have led to the development of numerous applications for music performance. Tablets are sufficiently powerful to run audio applications, and relatively affordable. Coupled with the immediacy of multi-touch, they can offer simultaneous interaction capabilities to musicians that go beyond what a laptop with a trackpad can offer. Although touch screen controllers for music such as the JazzMutant Lemur predate the iPad, the affordability of the iPad, and its processing power for audio made it a popular choice for musicians since its launch. Lemur itself has been ported to iPad (https://liine.net/en/products/lemur/), where it competes with other popular touch screen controller apps such as Touch OSC (http://hexler.net/software/touchosc). Reactable, a self-contained instrument, is another notable tangible music tool to have been ported to iPad, among other devices (http://reactable.com/mobile/). The field of tools for audiovisual performance targeting tablets is less varied than the field of tools for music performance. Few apps (tools or art pieces) for tablets aim to be used in a performance context for both audio and video output. Notable exceptions are Variant and its predecessor Thicket (http://intervalstudios.com), and Takete (http://refinedstochastic.com/takete.php). Due to the portability, expressive capabilities of multi-touch, and processing power, we have identified potential for tablets to be used as stand-alone tools for audiovisual performance, and not simply as controllers for other devices. To explore this potential, we have developed AV Zones.

Previous Work

In a preliminary, ethnomethodological phase, reported in Correia and Tanaka (2014), we conducted interviews with 12 audiovisual performers, asking them about their practice, the creative tools they use, their needs and desires as performers. This brought forth a series of key issues we retained as important for live audiovisual performance: modularity, flexibility and reconfigurability; ease of hardware/software integration; instrument-like expressivity and fluidity; integration of environmental elements; generative capabilities and diversity; communication of process to the audience; reliability and speed. The ideas from the interviews then informed a brainstorming workshop, with 19 participants. The five breakout groups produced five sketches of procedural audiovisual performance tools. Two were particularly successful in addressing the challenges set out in the workshop. Both rely on the expressive potential of multi-touch interaction, employing different solutions for reconfigurability: the former allows for loading and manipulating vector graphics, and the latter adopts a simplified data-flow mechanism for customization. The key issues identified from the interviews and sketches from the workshop influenced the design and development of AV Zones.

AV Zones

AV Zones (AudioVisual Zones) is an iPad app for audiovisual performance, composed of an audio sequencer/looper with a visualizer. It explores the interactive potential of a touch screen tablet for integrated musical and visual expression. By default, 3 audiovisual columns or "zones" allow for the manipulation of 3 audio loops. These zones are metaphorical adaptations of channels in a standard audio mixer. Each zone had 3 XY pads for audio manipulation: pitch shift, delay and filter. Each zone has its own sequencer as well. A visualization of each sound is overlaid on to the respective zone. There are 9 sounds available per zone, which can be replaced. Performing different gestures on each XY pad creates different results. The application is scalable: the number of zones, XY pads and sounds can be modified in the code. In a performance, only the iPad is used for audiovisuals: the visuals from the iPad are projected behind the performer, and the sound comes from the iPad as well. What the performer sees is also what is projected on the screen. The interface is shown in the screen, allowing the audience to better understand the performer's actions (figure 1; video: https://vimeo.com/144976072). AV Zones has been performed at: MonoShop opening, Berlin (May 2015); EAVI XIII, Amersham Arms, London (October 2015); VJ London, Juno, London (December 2015); and Seeing Sound, Bath (April 2016). AV Zones is open source and work in progress, built with openFrameworks and Maximilian add-on. The app is still being finalized, and will be submitted to the App Store within a few months. Meanwhile, the code is available on GitHub, and the app can be side-loaded manually on an iPad using Xcode 7 (https://github.com/AVUIs/AVZones).

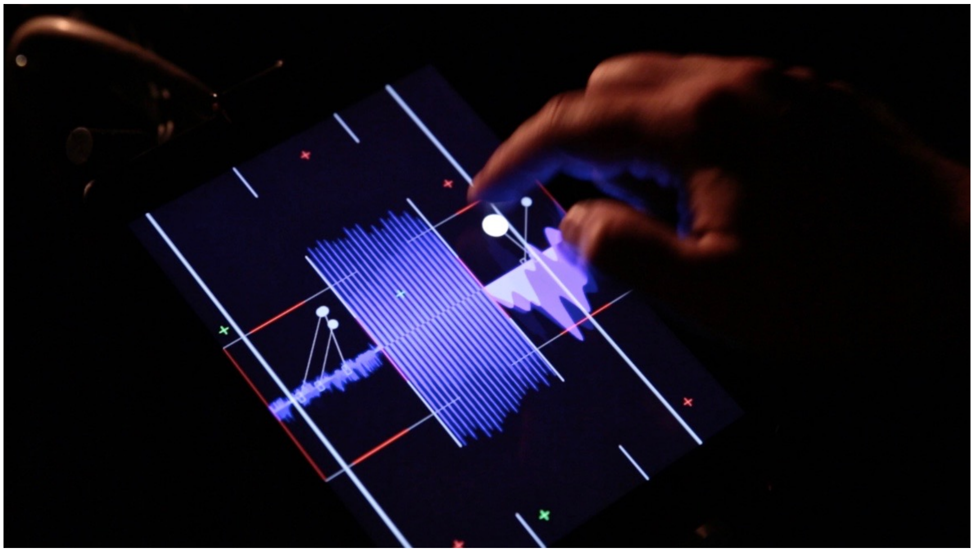


Figure 1. AV Zones

Biography

Nuno Correia is a researcher, media artist and musician. He is interested in enabling interactive multi-sensorial experiences. Since 2000, he has been teaching and conducting research in media art and design, in universities in Portugal, Finland, Estonia and the UK. Nuno holds a Doctor of Arts degree in new media from Aalto University (Media Lab Helsinki), with the thesis "Interactive Audiovisual Objects", and an MSc in innovation management from the University of Lisbon. Currently, he is a researcher at Goldsmiths, University of London (EAVI group), working on the project "Enabling Audiovisual User Interfaces". Nuno's work has been presented in more than 20 countries.

References

Correia, Nuno N. and Tanaka, Atau. 2014. User-Centered Design of a Tool for Interactive Computer-Generated Audiovisuals. *Proc. 2nd ICLI*. Lisbon.

Prototyping Audiovisual Performance Tools: A Hackathon Approach

Nuno N. Correia
Department of Computing
Goldsmiths, University of London
SE14 6NW London UK
n.correia@gold.ac.uk

Atau Tanaka
Department of Computing
Goldsmiths, University of London
SE14 6NW London UK
a.tanaka@gold.ac.uk

ABSTRACT

We present a user-centered approach for prototyping tools for performance with procedural sound and graphics, based on a hackathon. We also present the resulting prototypes. These prototypes respond to a challenge originating from earlier stages of the research: to combine ease-of-use with expressiveness and visibility of interaction in tools for audiovisual performance. We aimed to convert sketches, resulting from an earlier brainstorming session, into functional prototypes in a short period of time. The outcomes include open-source software base released online. The conclusions reflect on the methodology adopted and the effectiveness of the prototypes.

Author Keywords

Audiovisuals, performance tools, user-centered design, hackathon, expressive interaction, human-computer interaction

ACM Classification

H.5.1 [Information Interfaces and Presentation] Multimedia Information Systems–Animations, H.5.2 [Information Interfaces and Presentation] User Interfaces–User-centered design, H.5.5 [Information Interfaces and Presentation] Sound and Music Computing.

1. INTRODUCTION

We are interested in researching the development of tools for audiovisual (AV) performance in order to address the current practices and growing attention given to the field. This growth can be demonstrated by the several new festivals, seminars and publications (for example [4], [7], [10]) focusing in this area in recent years. In this study, taking into account needs identified in a previous stage [2], we aim to develop prototypes of tools for computer-generated audiovisuals, combining expressiveness, ease of use and visibility of interaction to the audience. In this paper, we present a prototyping approach based on a hackathon, and the resulting five prototypes.

2. TOOLS FOR AV PERFORMANCE

There has been an increased interest in different forms of "screen-based performance", adopting "a long litany of names such as audiovisual performance, real-time video, live cinema, performance cinema, and VJ culture" [11]. Three notable examples of contemporary audiovisual artists using computer-generated graphics and sound are Golan Levin, Toshio Iwai and Thor Magnusson. They are relevant to this study because they are concerned with creating interfaces and instruments for audiovisual expression. Levin developed a suite of works under the name *Audiovisual Environment Suite* (*AVES*) and described his approach to audiovisual performance as being based on painterly interfaces [6]. Iwai creates playful pieces, crossing genres between game, installation, performance (with works such as *Elektroplankton* and *Composition on the Table*) and audiovisual instrument (with *Tenori-On*) [9]. Magnusson uses unconventional Graphical User Interfaces (GUIs) and "abstract objects that move, rotate, blink/bang or interact" to represent musical structures [8]. The tools he develops are often made available online.

Most ready-made commercial software tools for live visuals (such as Modul8 or Resolume) focus on video playback and manipulation. Therefore, artists interested in using video for their performances have a choice of using either turnkey (and easier to use) software, or programming languages / environments (with a steeper learning curve, but offering more flexibility). For artists dealing with procedural graphics, however, there is a scarcity of ready-made, easy to use software.

The notion of expressiveness in AV performances deserves further attention. Hook et al. have studied VJ performances from an interaction design perspective [5]. Their key insights are grouped in terms of expressive interaction in VJ performances around three thematic categories: ***aspirational***, ***live*** and ***interaction***. From these, we will concentrate on the last one, as it relates the most to tool design. Within ***interaction***, Hook et al. identify the following key insights: *constraining interactions* (the importance of constrains and focus); *haptically direct* (using hardware controllers and having a physical connection to the system); *parallel interaction* (simultaneous control of multiple parameters); *immediacy* (immediate response from the software); *manipulable media* (desire for powerful and varied manipulation of media); *reconfigurable interfaces* (the ability to reorganize the controls to fit a particular performance); and *visible interaction* (to make the performer's interaction visible to the audience). The last three insights in particular relate to our aims of combining expressiveness, ease of use and visibility of interaction. Hook et al. focus on video manipulation, but state the need for tools creating procedural content, namely visual "devices that mimicked audio-synthesizers". Their study identifies key themes and needs, and points toward future paths, but does not produce new concepts or prototypes. We aim to further explore the paths laid out by their study, and produce prototypes as outcomes of that exploration.

3. HACKATHON ON TOOLS FOR AV

3.1 Hackathons and User-Centered Design

This study follows a User-Centered Design (UCD) approach. UCD is "a broad term to describe design processes in which end-users influence how a design takes shape" [1]. In particular, our approach is based on a hackathon, and the users are artists in addition to being computer programmers.

According to the online Oxford English Dictionary, hackathon is "an event, typically lasting several days, in which a large number of people meet to engage in collaborative computer programming". Hackathons share resemblances to workshops – "collaborative design events providing a participatory and equal arena for sharing perspectives, forming visions and creating new solutions" [12] – which have often been used in UCD studies. However, hackathons are more specific events, and require a technical skillset. Due to this requirement, hackathons are adequate for the creation of working prototypes. Workshops have been used as research methodology for modifying ("hacking") musical instruments [13], and some collectives regularly organize events dedicated to music or video hacking (for example, the London Video Hackspace[1]). Some organizations promote longer term do-it-yourself workshops to develop interactive projects (for example, MediaLab-Prado with the *Interactivos?* programme [3]). But the approach of using a hackathon as a UCD method for prototyping software for audiovisual performance is novel.

3.2 Preparing and Running the Hackathon

In a preparatory phase (presented in [2]) to the research reported here, we conducted interviews with 12 audiovisual performers. We asked the artists about their practice, the tools they use, and their needs as performers. This generated a series of key ideas that informed a brainstorming workshop. The outcomes of the workshop were sketches of tools for audiovisual performance using procedural sound and graphics. These sketches were the input to the hackathon.

A call for the hackathon was distributed among the Goldsmiths and London Video Hackspace communities. Previously to the hackathon, we interviewed four of the more experienced participants to validate our plans for the event, and obtain suggestions. The two-day (8 hours per day) hackathon took place in December 2014, and a pilot was conducted exclusively with Goldsmiths students. 18 participants took part in the hackathon, and five in the pilot (five female and 18 male in total). The participants were divided into six groups, taking into account a previously filled-in questionnaire identifying preferred programming languages and development environments. Groups were created focusing on Processing (three groups), openFrameworks (two) and Cinder (one). Participants with an emphasis on sound (Max/MSP and Pure Data) were distributed through those groups.

The hackathon started with a presentation on the previous stages of the study (interviews and workshop) and results achieved so far. The sketches produced in the workshop were presented in detail, and participants were invited to adopt features of the sketches into their projects. The structure for the workshop was outlined. The first stage was the development of sketches, followed by software programming. A follow-up one-day event was agreed for finishing the projects. One of the groups did not continue to the follow-up event, therefore there were five resulting projects in total.

We used GitHub, a web-based Git repository hosting service, to manage and distribute the code created for each project. We created an "organization" page within GitHub[2], and each group created a code repository within it. GitHub facilitates the identification of contributions, and modification ("hacking") of software by others. The groups were encouraged to release their projects as open-source.

4. RESULTS

The resulting five projects from the hackathon are: *ABP*, *drawSynth*, *Esoterion Universe*, *GS.avi* and *Modulant* (Figure 1). They are available for download or for code modification, from the respective GitHub pages. A website was created for documenting the projects.[3]

ABP is an animation engine and sound visualizer where the user can define color, geometry and animation parameters (position, rotation, size, motion vector) with a GUI. The GUI can be shown to the audience. The visuals are based on a particle system, which is sound-reactive. The visual module is built with Cinder and the sound module is created with Pure Data. The sound consists of two drum synthesizers. Information is communicated between the two programs using the OSC (Open Sound Control) protocol.

drawSynth consists of a GUI to control sound and image. Users can draw vector shapes and select colors. By doing this, they control the FM synthesis engine. The position of the vector points on the screen affects parameters of the synthesis, such as number of carriers, modulators and their frequency. The project is built with openFrameworks for graphics and interaction, and the Maximilian openFrameworks add-on is used for sound.

Esoterion Universe starts with an empty 3D space that can be filled with planet-like audiovisual objects. The objects can be manipulated and can be given different appearances and sounds. The visual component of the objects is audio-reactive. Users can navigate in space and the audiovisual outcome is influenced by that navigation. Generic, media neutral terms such as warmth, sharpness, size and roughness are used to characterize and connect sound and visuals. This semantic approach was chosen instead of a one-to-one parameter mapping. The GUI consists of sliders distributed concentrically, in the shape of a star graph, embedded in the center of the object, and integrating aesthetically with the objects. openFrameworks with openGL is used for graphics and interaction. The sound component is a granular synthesizer built with Max/MSP. OSC is used for communication.

GS.avi is a gestural instrument that generates continuous spatial visualizations and music from the input of a performer. The features extracted from a performer's gesture, using the GVF software[4], defines the color, position, form and orientation of a 3-dimensional Delaunay mesh – its composite triangles, vertices, edges and walk. The music, composed using granular synthesis, is generated from features extracted from the mesh – its colors, strokes, position, orientation and patterns. The project was created using Processing and Max/MSP. OSC is used to communicate between the two.

Modulant allows for the creation of images and their sonification. The present implementation is built upon image-importing and freehand-drawing modules that may be used to create arbitrary visual scenes, with more constrained functional and typographical modules in development. The audio engine is inspired by a 1940's synthesizer, the ANS, which scans across images. In this scanning, one axis is time and the other axis is frequency. *Modulant* thus becomes a graphical space to be explored sonically and vice-versa. The project is built with Processing for graphics and interaction, and Ruby with Pure Data for sound.

A public performance took place in February 2015[5], and an audience evaluation was conducted. An additional evaluation of the projects by AV performers has also taken place meanwhile. We are now in the process of analyzing the results from these evaluations.

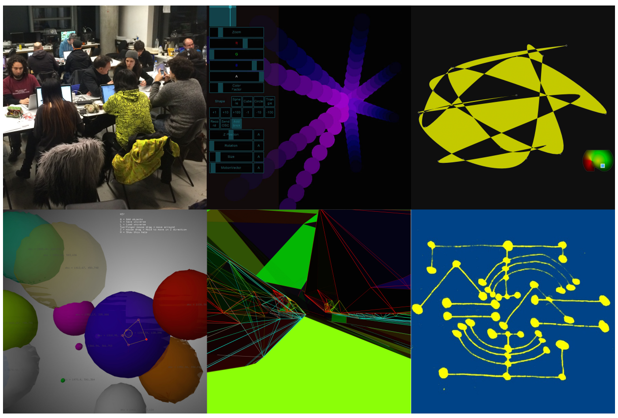


Fig. 1 Left to right, top row: hackathon, *ABP*, *drawSynth*; bottom row: *Esoterion Universe*; *GS.avi*; *Modulant*.

[1] http://www.videohackspace.com
[2] https://github.com/AVUIs
[3] http://avuis.goldsmithsdigital.com/gen-av-feb-2015/

[4] http://eavi.goldsmithsdigital.com/resources/gesture-variation-follower-gvf/
[5] Video from the performance: https://vimeo.com/124065089

5. DISCUSSION

5.1 Prototypes

Looking at the five projects through the lens of expressive interaction of Hook et al., the projects diverge in their degree of conformity with the design insights, particularly in the thematic category ***interaction***, and the related key themes identified for this study of *manipulable media*, *reconfigurable interfaces* and *visible interaction*.

Two of the projects offer immediate *manipulable media* capabilities in the visual domain – *drawSynth* and *Modulant* – since they are drawing applications. They allow for varied graphical manipulation: vector-based, in the case of *drawSynth*, and bitmap-based, in the case of *Modulant*. Of the two, *Modulant* is the most versatile, since it provides a larger diversity of drawing tools, and it allows for the saving and loading of bitmap images. *drawSynth* is the most fluid, since it produces immediate sonic results upon drawing. However, both offer very simplistic drawing capabilities at this stage. The graphics of the remaining projects are not as manipulable.

Sonically, the projects that offer greater manipulation are the ones based on granular synthesis – *Esoterion Universe* and *GS.avi* – as they allow for the loading of different sounds, leading to more varied results. *Esoterion Universe* offers more sonic manipulation capabilities, based on its semantic approach, although they do not change the sound characteristics drastically. The sonic manipulation capabilities of the different projects are not very powerful, leading to results that are not substantially varied.

None of the projects allow, on the surface, for truly *reconfigurable interfaces*. However, the fact that the projects are open-source allows for reconfiguration. The reconfiguration of interaction is particularly easy in the case of the projects that use the OSC protocol for data communication: *ABP*, *Esoterion Universe* and *GS.avi*. The OSC messages can be easily rerouted to other parameters, or entirely different sound generation sources.

Regarding *visible interaction*, most of the projects (*ABP*, *drawSynth*, *Esoterion Universe* and *Modulant*) display the media manipulation act to the audience. The screen seen by the performer is also intended to be the screen shown to the audience. *ABP* and *Modulant* rely on more traditional GUIs – buttons and faders in the former, and a "drawing tools palette" in the latter. In *drawSynth*, the drawing tool is fixed – a line-drawing tool, complemented by a color picker. In *Esoterion Universe*, a radial GUI with faders is embedded in the "planets". In all four, the cursor is showcased to the audience, revealing the choices made. In *APB,* both GUI and cursor can be hidden. *GS.avi* is the exception – it was designed for gestural interaction, preferably with a tablet. The interaction with the project is not made visible to the audience. All five projects rely in keyboard shortcuts, in alternative to, or in addition to, the pointer- or touchscreen-based interaction.

5.2 Hackathon Approach

The hackathon approach was successful in converting sketches into functional prototypes in a short period of time. Adding an extra day to the planned two-day hackathon proved to be decisive for the completion of projects. Therefore, one important lesson learnt was to have a flexible timeline for the hackathon. However, an extension of the hackathon has the risk of leading to a higher number of dropouts – some of the participants did not continue to the extra day.

GitHub proved to be an essential tool for groups to collaborate and share code, not only during the hackathon but also for collaborating remotely outside of the event. Additionally, it provided an important platform for sharing the projects and their code. One of the groups used GitHub for communication and knowledge sharing as well.

Organizing the teams around the same technology was important for the success of the hackathon. Running a questionnaire before the workshop for identification of preferred technologies saved valuable time during the event. However, an AV project does not have to be built using a single technology. In particular, the sonic and visual components can be built with different technologies, and communication between those can be facilitated by OSC. Building the project with a single technology does have an advantage – ease of distributing and setting up the software.

6. CONCLUSIONS

Some of the prototypes (notably *Esoterion Universe* and *Modulant*) present potential for addressing the needs of audiovisual performers. As prototypes, all of the projects have room for improvement. But even at this stage, they contain strengths that can be of use and inspiration to other projects. Additionally, the code from these prototypes can be reused, given their release as open-source. The GitHub platform, where the projects are hosted, facilitates this reuse. Most of the projects adopt the OSC protocol, allowing for easy re-mapping of audio and visual parameters. Hence, these hackathon outcomes are also *hackable instruments* that allow for "discovery of novel working configurations", as defined in [13].

With the present study, we introduced a novel approach for the user-centered prototyping of tools for audiovisual performance, based on a hackathon. This approach was successful in converting sketches into prototypes in a short period of time. It was also successful in providing different creative perspectives to an initial challenge. We believe that hackathons offer potential to be used for UCD studies in collaborative prototyping for creative fields beyond audiovisual performance. We are planning further developments taking into account the forthcoming evaluation results.

7. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n° 627922. We would like to thank the hackathon participants, Alessandro Altavilla for the photo and video documentation, and Peter Mackenzie for technical assistance.

8. REFERENCES

[1] Abras, C., Maloney-Krichmar, D., and Preece, J. User-Centered Design. In W. Bainbridge, ed., *Encyclopedia of Human-Computer Interaction*. Sage Publications, Thousand Oaks, 2004.

[2] Correia, N.N. and Tanaka, A. User-Centered Design of a Tool for Interactive Computer-Generated Audiovisuals. *Proc. 2nd ICLI*, 2014.

[3] Corsín Jiménez, A. and Estalella, A. The hospitable prototype. *Prototyping cultures*, 2010.

[4] Faulkner/D-Fuse, M. *VJ: Audio-Visual Art and VJ Culture.* Laurence King, London, 2006.

[5] Hook, J., Green, D., McCarthy, J., Taylor, S., Wright, P., and Olivier, P. A VJ centered exploration of expressive interaction. *Proc. CHI '11*, ACM, 2011.

[6] Levin, G. A Personal Chronology of Audiovisual Systems Research - Golan Levin and Collaborators. *Proceedings of NIME '05*, (2005).

[7] Lund, C. and Lund, H. *Audio Visual: on Visual and Related Media.* Arnoldsche, Stuttgart, 2009.

[8] Magnusson, T. Screen-Based Musical Instruments as Semiotic Machines. *Proc. of NIME '06*, 2006.

[9] Nagle, P. Yamaha Tenori-On Review. *Sound On Sound*, February 2008.

[10] Rainer, C., Rolling, S., Daniels, D., and Ammer, M., eds. *See This Sound.* Walther Konig, Köln, 2010.

[11] Salter, C. *Entangled: Technology and the Transformation of Performance.* MIT Press, Massachusetts, 2010.

[12] Soini, K. and Pirinen, A. Workshops - Collaborative Arena for Generative Research. *Proc. DPPI '05*, 2005.

[13] Zappi, V. and McPherson, A. Design and Use of a Hackable Digital Instrument. *Proc. 2nd ICLI*, 2014.