Sound Matters: a framework for the creative use and re-use of sound:

Field recordings and Speech

FIRST STAGE REPORT

Funded by: JISC – Research Data Spring

Funded amount: £10000

Team

PI: Dr. Ximena Alarcon, Research Fellow CRiSAP, LCC – University of the Arts London

Co-I: Prof. Cathy Lane, Co-director CRiSAP, LCC – University of the Arts London

Administrator/Communications support: Lisa Hall, CRiSAP, LCC - University of the Arts London

Project's websites:

http://soundmattersframework.wordpress.com

http://www.crisap.org/index.php?sound-matters-a-creative-framework

Wiki: https://wiki.research.data.ac.uk/Sound_Matters_Framework

Twitter: #soundmatters

Project's Aim

To create a framework for the use and re-use of sound that facilitates the interrogation and relational playback of non-musical sound material specifically **field recordings** and **speech**: A Listening-Led Environment.

Objectives

1. To create an interdisciplinary community for the use and re-use of digital sound in meaningful and creative ways, expanding the potential outreach and value of sound material in its own terms.
2. To create a Listening-Led Environment that offers accessible interfaces and tools that links the processes of relational playback and interrogation to be used creatively by an interdisciplinary community

This report contains the development and findings of the project Sound Matters Framework in its First Stage, and the proposal for the Second Stage.

1. First Stage Aim

The first stage of the research aimed to identify common issues (difficulties and achievements), needs and coping strategies shared by interdisciplinary researchers in the UK and abroad who are researching with sound, and to envision together a technical framework that can be used as a reference to facilitate their research using open source and user-friendly accessible technologies.

Specific Activities:

- Activate a community of users through: a. interviews; b. online gathering; c. Co-design workshop
- Conduct a Technical Review
- Expand the framework description (uses and technologies)
- Write a proposal for the second stage

2. Activities

2.1 Creation of a Community

2.1.1 Interviews-starting the community:

20 Researchers/Artists[1] working with Field Recordings, Speech and Archives were interviewed about their research and creative processes, goals, human and technical issues involved, within the framework.

2.1.2 Online Gathering:

Derived from the interviews a framework was expanded with specificities about each process: *Interrogation* and *Relational Playback*, and it is live on a virtual site Mural.ly[2] for comments. The framework describes existing tools and methods of *Interrogation* of Field recordings and Speech, and the use of research processes, and of composer and users' interfaces for *Relational Playback*. Also it describes the different types of archives used by people on their creative and experimental research.

The interviews nourished the framework, and led to the publication of nine representative stories in the blog of the project[3]. The stories are divided into Field Recording, Speech and Archive stories. The idea with these publications is to open up all the aspects of the process which are not only subscribed to technical processes. These are open for the general public to comment on.
The activity of the blog, was promoted through a Facebook event page, twitter feeds, the CRiSAP website, and CRiSAP, UAL email lists and LCC newsletter. This communication strategy was led by the Administrator Lisa Hall.

2.1.3 Co-design workshop:

- 16 Researchers/Artists working with field-recordings, speech and archives attended to the co-design workshop led by the PI. The objective of the workshop was to gather the community in a closer manner to create conversations about the project, and creatively think together on methods and processes mediated by technology within the focus of the research.
- In the workshop the PI offered an update of the research and its findings, and invited two guest speakers (one hour per speaker) to stimulate thinking and discussion about: 1. Archives (Andrea Zarza, Independent Researcher and Curator British Library), and 2. Algorithmic Tools (Ron Herrema, Algorithmic composer).
- At midday participants were able to experience the latest technologies for Search and Retrieval and Relational Playback, as exhibited by:
  Chris Baume (PhD Student University of Surrey): BBC's new speech editing Prototype that features speech-to-text and speaker diarization, and the BBC World Service Archive which includes topic identification, speaker identification and Crowdsourcing features.
  Cathy Lane (Co-director CRiSAP): BEAM installation artwork, which uses a Complex Max/MSP patch using data from the coming and goings around the port of Kochi in South India as parameters to trigger sounds associated with maritime life and the spice trade in the area.
  Ximena Alarcón (PI): Exhibiting Online Tools from labs freesound as developed by the Music Technology Group at the University Pompeu Fabra.
- In the afternoon participants divided into three groups designed case studies, imagining contexts, processes and interfaces that involved working with a

collection of sounds and the processes they will perform with it, as well as the interfaces. These designs led to some discussions and conclusions of the day: opening up to materials that are not necessarily sonic; working with algorithms that work via associations of different types; allowing of participants in a research process to create their own tagging systems (Fig 1, Fig 2, Fig 3).

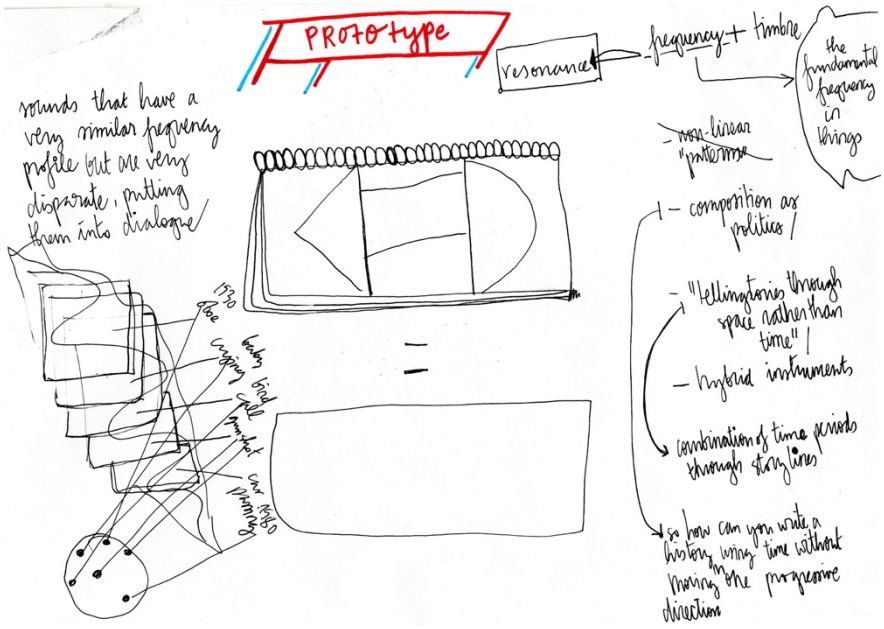

Fig. 1 Exercise with participants in co-design workshop Group 1

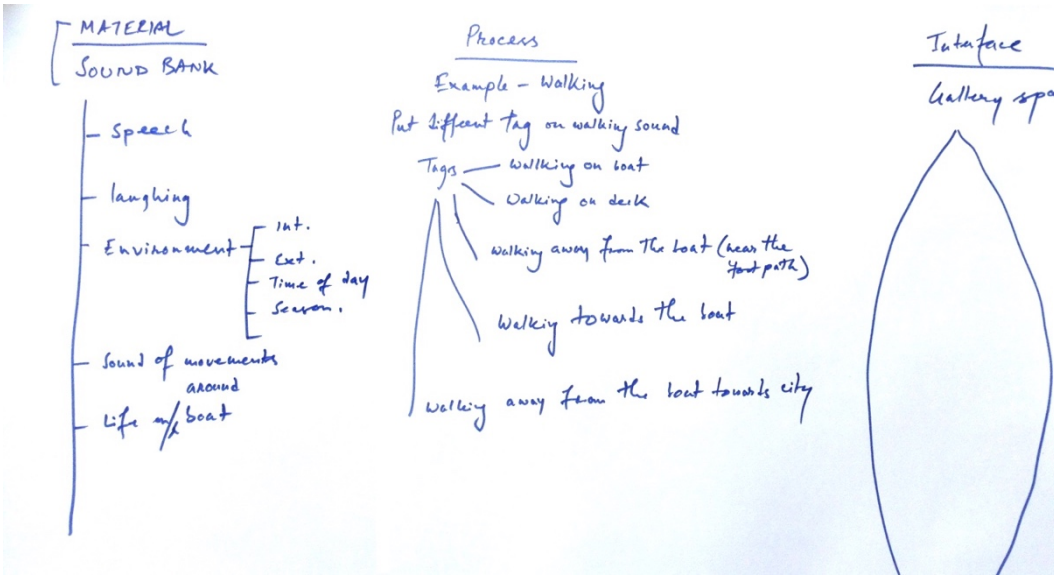

Fig. 2 Exercise with participants in co-design workshop Group 2

Fig. 3 Exercise with participants in co-design workshop Group 3

- A 6'31" video (Fig 4) documenting the co-design workshop and offering a background of the project has been produced and can be accessed on https://vimeo.com/133219234

Fig. 4 Video documenting Sound Matters community building process

2.2 Technical Review and Context

2.2.1 Technical Review:

- A Technical Review, via Literature Review and direct communication with developers and researchers[4], was conducted by the PI about the current options for *Interrogation* and *Relational Playback* of sound material: Creation of meta-data, Search and Retrieval meta-data, and Sonic Information Retrieval, and received collaboration from University Pompeu Fabra, University of Surrey, Queen Mary University of London, and Goldsmiths.
- The most up to date technologies for Search and Retrieval of Speech were explained by Chris Baume, as developed within his PhD research in Surrey U and for his work in the BBC, speech to text recognition. In terms of Sonic Information Retrieval involving a variety of sounds, the most advanced systems were demonstrated by Frederic Font, Research Fellow and leader of the freesound project at the University Pompeu Fabra in Barcelona; this involves the development of the freesound API (which works with tag recognition) and the use of the Essentia Library (for sonic information retrieval).
- The PI attended to SONAR Festival in Barcelona to witness the music hack day, organised by the University Pompeu Fabra, where the most advanced libraries and techniques for Search and Retrieval were used. Also she attended to workshops, talks and the SONAR Market, where she networked and talk to programmers and developers in various fields about accessing to big libraries of sound creatively. This visit opened up the thinking of diverse interfaces that are being created from modular systems to allow more people to create from the front-end perspective, with more accessible ways to understand processes happening in the back end.
- PI and Co-I attended to the Digital Conversations event hosted by the British Library about experiences on Digital Music Retrieval. This opened up the conversation with researchers at Surrey University Prof. Mark Plumbley, and Welcome Trust Fellow Dr. Erinma Ochu.
- Support from sound technical tools development has been received from the Bergen Center for Electronic Arts (Norway), specifically looking at their project *Jamoma Modular*[5]. This collaboration is born from the need of prototyping with modular systems that can be used for a variety of interfaces for the members of the community.

- Support from Grassroots archival system, involving Ethical issues and Copyright has been received from the Center for Digital Archeology (California), specifically looking at their project *Mukurtu*[6].

2.2.2 Context Support:

Support about the state of sound archives, was received from Luke McKernan, Lead Curator, News and Moving Image from the British Library, and leader of the AHRC funded 'Opening up speech archive project', conducted two years ago, and from James Knight from 'Save our Sounds' project. Also context about local sound archives accessibility from Siân Mogridge from Hackney Archives.

2.3 Scholarly Dissemination:

An abstract about the project was sent to the international conference Sound, Image and Data, 2015 to be held at New York University, and it was accepted. PI will attend from 23rd to 25th of July to present the first findings of the project. It would be also an opportunity to know more technical and artistic context.

3. Main Findings:

3.1. Interdisciplinary Goal

The main interdisciplinary goal perceived from the interviews and research is the making and re-making of memory, re-writing of history through sound, traveling though the collections/archives/databases of sounds in time and space.

- Sound Matters (SM) is being configured as a conceptual, methodological and technological framework, from the perspective of artists and interdisciplinary researchers (Fig. 5 and Fig. 6).

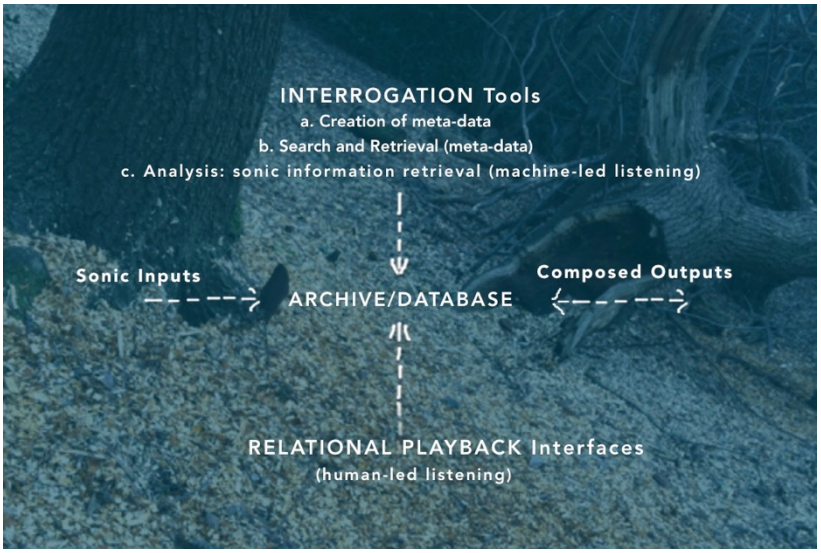

Fig. 5 Initial proposed framework

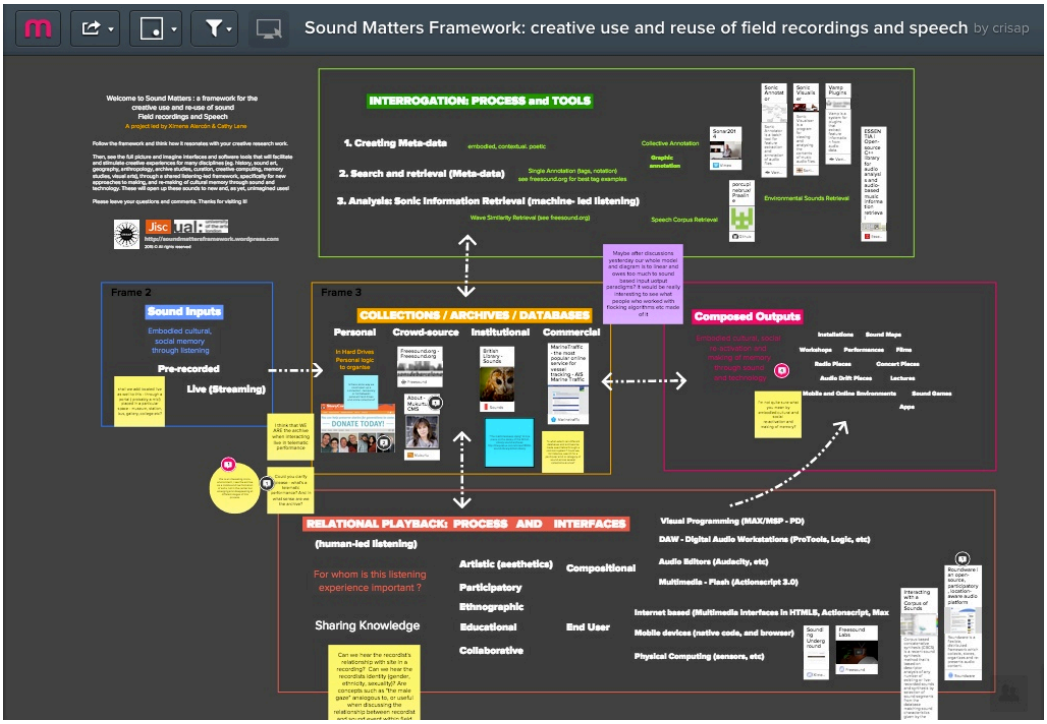

Fig. 6 Extended Framework at Mural.ly for comments and input from Sound Matters community

3.2. Framework elements

3.2.1 Input

Most of the interviewees record their own material. The act of recording is part of framing a research or artwork. Some of the recordings are made in collaboration with people from communities living in a particular place. Some interact with archives (not only sonic). Some work with live streaming. Some interactive work they make uses voice/ speech as being recalled by listening to field recordings; these voices become part of the artwork. The distinction between field recordings and speech is not that strong (Fig 7).

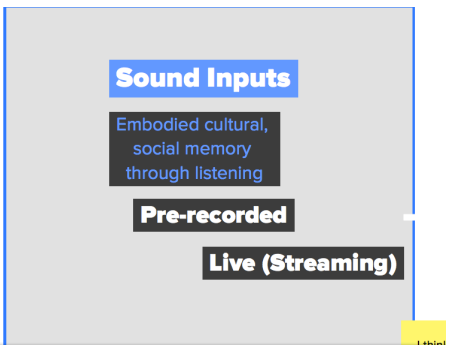

Fig. 7 Input at Mural.ly

3.2.2 Archives

There is a general preoccupation for Sound Archives: Save our Sounds. Creating personal collections, residing in hard drives. Some researchers question about the accumulation of data and include in their practices the deletion of recorded material. It seems that sharing sounds used in their research seems more relevant if sharing with their peers, or with the communities they have worked with. Archiving sounds in British Library seems interesting but very demanding. There is also interest of archiving their material in places such as Archive.org. Dedicated to Crowd-Sourcing sound is the site *Freesound*[7], which has developed software for retrieval, tagging, derived of many years of research at the University Pompeu Fabra. There exist social data repositories, and we highlight *Mukurtu* as grassroots based space that attend the needs of ethnographic research offering a safe space to manage all kind of material for different community projects. It has been developed after years of research in the Center for Digital Archeology in CA – US. It uses innovative copyrights licenses such as TK (traditional knowledge), and the option of deletion when the owner of the file wishes to do so. They also have developed a Mukurtu mobile app. Researchers/Artists also use non-sonic databases for the creative interaction with sonic material, adding data flow context of social activities (Fig.8).

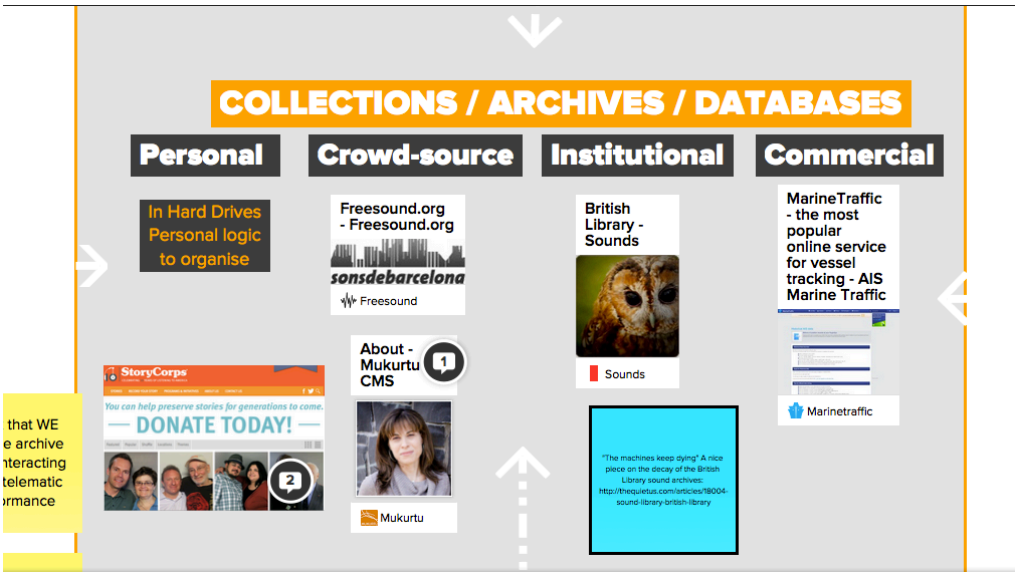

Fig. 8 Collections/Archives/Databases in Mural.ly

3.2.3 Relational Playback

Relational Playback (Fig. 9) is so far the most interesting process for the Sound Matters interviewees. The term has different understandings:
It is the most important process for ethnography, e.g. playback sound material to people in the field, opening space and paths for the process of recalling a place, and travelling through time. *Relational Playback* is also understood as compositional process, understanding and developing aesthetics and narratives. Playing back sound files from different sources, spaces, places and from different times creates different understandings of a determined cultural context. The experiences involve collaboration with artists, researchers and communities in experimental projects with educational outputs. This is in itself part of the research process involving risk, and experimentation. When involving communities, this process includes technologies from basic playback (hardware iPods) to sophisticated interfaces for installations (using Visual programming environments such as Max/MSP), and also online environments for end user: general public. Relational playback is also understood as a standard way of working on the studio with technologies such as sequencers, editors for the researcher/artist.

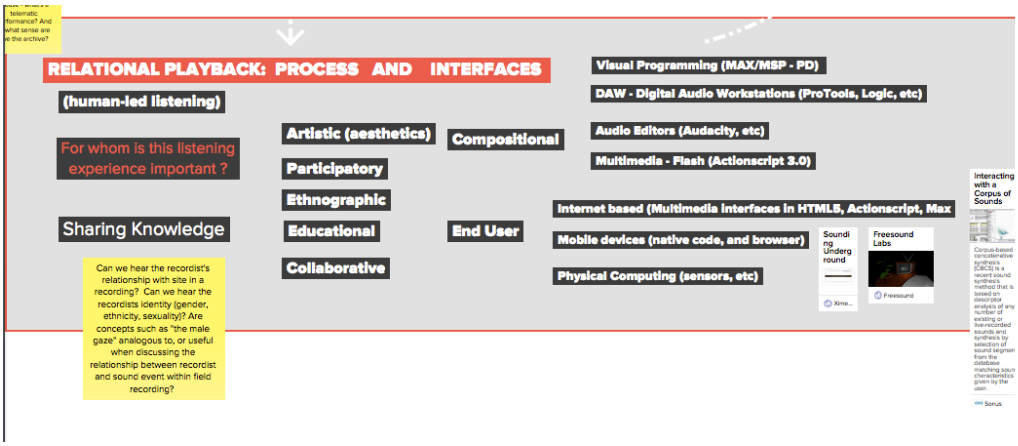

Fig. 9 Relational Playback: process and Interfaces, in Mural.ly

In our technical review we were envisioning a modular system that help to develop functions according to specifications of working with Field Recordings and speech, in the way in which Max/MSP visual programming environment offers. *Jamoma Modular*[8] (based on Max/MSP, which contains a growing number of modules dedicated to data, audio or video processing) seems the most appropriate software environment to link Relational Playback processes with archives and interrogation

tools. It consists of a set of modules that develop different audio processes, which can be linked to other modules. It has been used for installations and live performances. It has portability to work with C++, and to be viewed in a variety of interfaces.

3.2.4 Interrogation

This is an important process acknowledged by researchers but it not used too much, part because of lack of knowledge and technologies of how it might work, but also because as they know this process (tagging, categorisation), it is perceived as very demanding and outside the creative process. Researchers rely on their memory to find sounds within their collections and label their files by themes, date, place and some sonic characteristics. Some of them have critical perspectives about it: 'why cataloguing?' some of they question 'This is probably against flexible nature of sound'. Also some of them pointed out issues linked to the politics involved in machine processes, e.g. trust in the machine, or being part of a culture of surveillance. It seems clear from the conducted interviews that from an archivist's perspective, *Interrogation* is the most important process, while from the artists' and researchers' perspective: listening via relational playback is the most important process, listening without a goal.

This process incorporated alternative perspectives to understand metadata, as suggested by the Brazilian scholar Dr Carlos Falci: "a poetics of metadata can create archives able to combine invention and discovery, giving rise to imaginary places of memory."[9]

In an interview with the scholar[10] he suggested to include the "creation of meta-data" as part of the framework of Sound Matters. Also creative perspectives of annotation have been included as developed by Dr. Michael Gallagher[11], and the ones offered by Andrea Zarza in the co-design workshop, with the Sonic Time Capsule[12], which invite us to listen to the present and select the most important memories to keep. Thus, 'Creating meta-data' was incorporated as an element of the process of Interrogation, bringing the poetics and social context of it.

Single or collective annotation (feedback) has been developed by projects such as Music Circle – Goldsmiths. Music Circle uses feedback within a community, which can be explored, however feedback doesn't necessarily bring meta-data to play with

sounds later. The system is based on the sharing of tracks. Freesound tagging system, as developed at the University Pompeu Fabra, by researchers at the Music, Technology Group, is based on crowd-source tagging and has developed software to recognise certain tags and their associations, suggesting the user other tags; this system helps to make the search more accurate, and according to user needs (Fig.10).

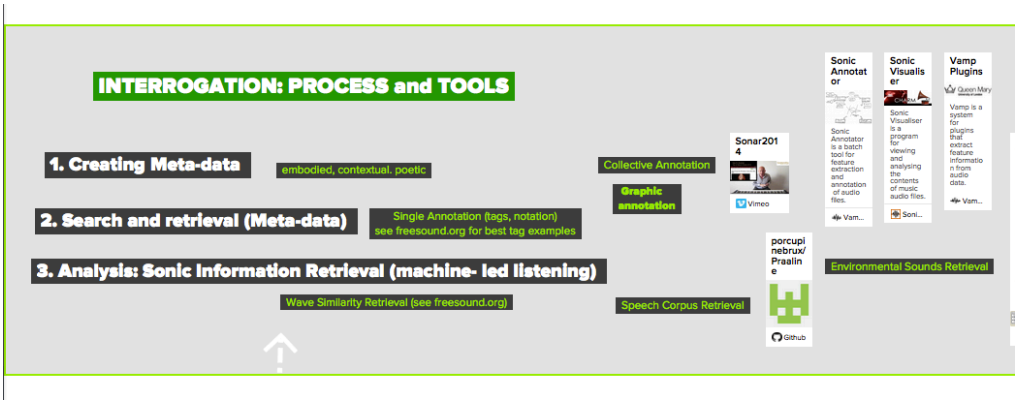

Fig. 10 Interrogation: Process and Tools, in Mural.ly

'Search and Retrieval' more up to date software tools identified were the Freesound API, which combines tag search with sonic information retrieval. This includes both field recordings and speech (spoken word and voice). Freesound has very effective tools but it is too massive, and can create a sense of being lost in the crowd. We need to see how it works for specific projects and communities with sensitive material. The premise of freesound is that all is free of use, using Creative Commons licensing. This might not be the most appropriate for ethnographic practices and the ethics of human research. We would like to use the freesound API link with other spaces/repositories such as Mukurtu.

On the other hand, we identified BBC Speech to Text and Diarization systems, which have been based on the hyperaud.io platform to navigate, search and edit spoken word. These seem to be mainly studio recordings, which acoustic is controlled for speech recognition.

'Sonic Information Retrieval' most up to date software is the Essentia library, developed at the Music Technology Group (MTG) at the University Pompeu Fabra (UPF). Freesound API uses Essentia[13], and any sound can be compared based on sound wave recognition and many other parameters such as Pitch, High Frequency

content, Beats Loudness, Onset rate etc (Fig 11).

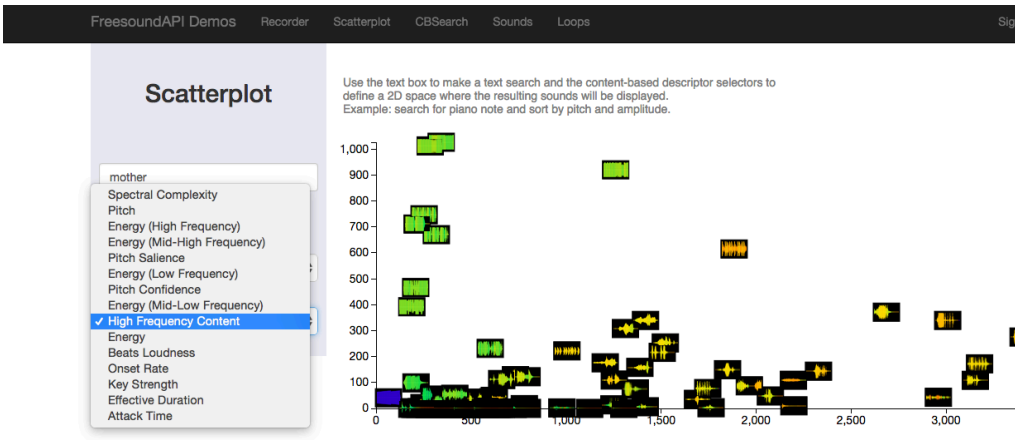

Fig. 11 Freesound API Demo.

Specific approaches to identify environmental sounds characteristics have been researched by the MTG at the UPF, using Gaver's taxonomy which looks at the acoustic properties of the sounds, based on Interacting materials such as: vibrating solids, aerodynamic Sounds, and liquid Sounds (Roma et al, 2010). This is an interesting research, and more discussion is needed with the community to know if this approach is relevant for their needs (Fig. 11).

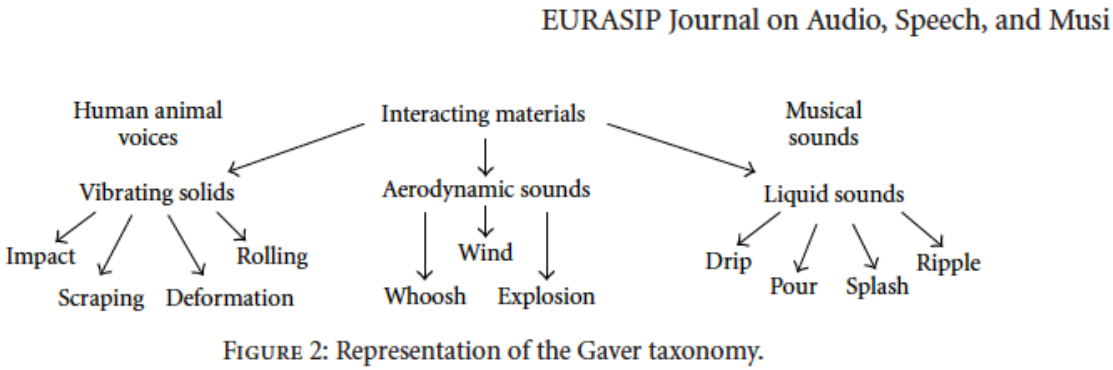

Fig. 12. Taken from "Ecological Acoustics Perspective for Content-based Retrieval of Environmental Sounds" (Roma et al, 2010)

3.2.5 Composed outputs

The composed outputs developed by members of the Sound Matters community are varied according to art forms: pieces, installations, mediated performances, films, computer games, internet-based environments, sound maps, mobile apps, and radio programs. Other outcomes are also workshops and lecturers, understood as creative spaces. (Fig. 13)

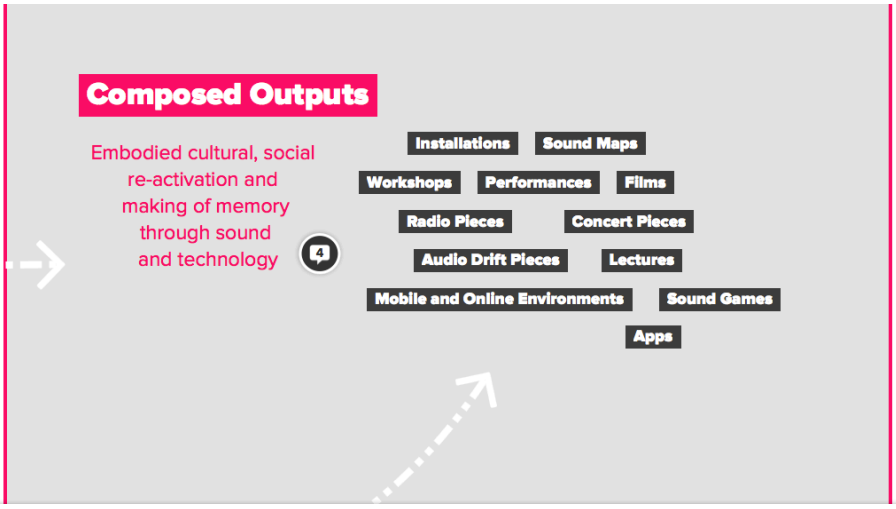

Fig. 13 Composed Outcomes, in Mural.ly

4. First Stage conclusions:

4.1. Archives

With the interviews and the co-design workshop, people became aware of the archives they are creating, and the existing tools and possibilities (e.g. algorithms) to access creatively to it.

Archives from the Sound Matters (SM) community are created from daily life of many communities who wouldn't ordinarily be in institutional archives. The aesthetics and output forms of the SM community are varied. On the other hand, it seems that there is not too much interest to 'deposit' their files in a crowd sourced environment (such as freesound), as many of the SM community projects rely on the dedication of a research/artistic project, and the context. Also this raises the idea that archives are not necessarily only the ones held by big institutions and these don't have to be necessarily deposited there.

It is important to make archives available by a selected community that understands the social value of it, and explore creativity in the processes of annotation, search and relational playback.

Thus, we propose to experiment working from scratch with an online open source dedicated site such as Mukurtu.org, which is open source and deals with social ethics and copyrights. It has been developed by the Center of Digital Archeology in the US. It is open source and can be open to interact with open source tools for *Interrogation* and *Relational Playback*.

4.2. Interrogation:

Creating meta-data

It is important to make archives annotated by the communities involved in the project, and not only by the recordist/researcher/artist.

Search and Retrieval

Being in touch with technologies for search and retrieval of large collections of sounds such as: speech-to-text to facilitate transcription; interfaces that search by a mixture of parameters combining tags and audio information retrieval; and interfaces that trigger sounds based on non-sonic metadata, was innovative for the participants in the co-design workshop.

Although people were enthusiastic about the speech-to text tool by Chris Baume, we need to find the creative side of this to be integrated on the Sound Matters (SM) listening-led environment. As it is, it follows the radiophonic model, and the original speech is made in a studio (controlled acoustic environment).

Thus, at the moment we think that the community will benefit of a tool for creation of meta-data, which links to own files/archives and perhaps other archives, and that encompasses a poetic and creative listening approach. A collective annotation space for people.

Equally the examples from freesound API are interesting mixing both parameters: tags and audio, and it would be great to know if we can adapt this for our environment.

4.3. Relational Playback

The introduction of the notion of algorithms and the creative possibilities thinking from this perspective opened perspectives for the community and three creative ideas were

born out of the exercise of thinking of context, process and interfaces.
This suggests the need of having modular structures with algorithmic parameters that the community can use to listen to archives in many different possibilities, traveling through time and space, and using this listening created environment to interplay with a diversity of interfaces.
BEAM Max patch that mixes non-sonic metadata with sound, was a very nice and complex example of it.

Thus, we propose to focus on the relational playback as a creative process, and to create a modular prototype using *Jamoma Modular*. The modules can be build according to the specifications of the SM community, and can grow. Jamoma Modular is open source, and in consultation with the developers of the other software involved in the framework, such as *Mukurtu*, *Essentia RT* Library and *Freesound API*, it is possible to make connections between all of these. All software is Open Source, and available in Github. All these software is supported by strong communities, which allow its sustainability.
For the second stage we propose to explore the connections between these software packages and create a prototype using Jamoma modular (Fig. 14):

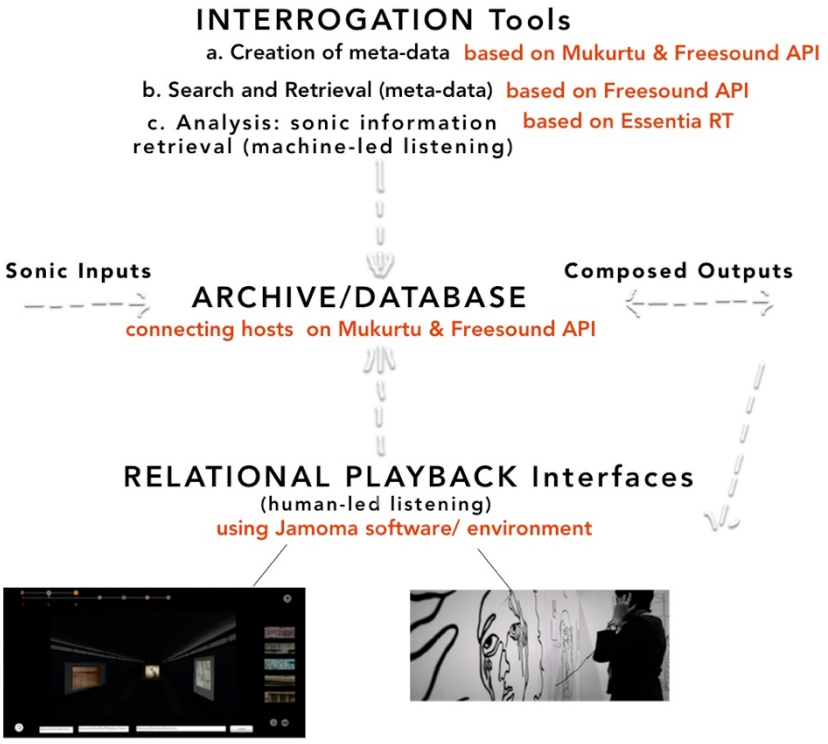

Fig. 14 Proposed Technical Tools for Prototype in Second Stage

[1] Interviewees: Isobel Anderson, Sound Artist and PhD student, Sonic Arts Research Centre, Queen's University Belfast; Catherine Clover, PhD student RMIT University Melbourne; Dr Rupert Cox, Senior Lecturer Social Anthropology, University of Manchester; Peter Cusack, Artist and musician, researcher from CRiSAP, London College of Communication, University of the Arts London. Favourite Sounds; Prof. John Drever, Acoustic Ecology and Sound Art, Goldsmiths, University of London; Dr Michael Gallagher, Musician, Lecturer and Researcher in Human Geography, Manchester Metropolitan University; Clay Gold, Sound Recording Artist; Dr. Ron Herrema, Composer, Developer and Generative Artist; Dr. Holly Ingleton, Cultural Worker and Feminist Sound Studies Scholar; Prof. Cathy Lane, Composer, Sound Artist and Researcher, CRiSAP, LCC University of the Arts London; Sian Mogridge, Hackney Archives; Sukanta Mujumdar, The Travelling Archive; Dr. Katharine Norman, Independent Researcher and Composer; Tullis Rennie, PhD Research Student, Sonic Arts Research Centre, Queen's University Belfast; Dr. Tom Rice, Lecturer in Anthropology, at University of Exeter; Dr. Adam Parkinson, Researcher at Embodied Audiovisual Interaction Group, Goldsmiths, University of London; Dan Scott, PhD student CRiSAP, UAL; Mark Peter Wright, Artist-Researcher, CRiSAP, LCC University of the Arts London; Andrea Zarza, Independent Researcher, Curator of World and Traditional Music, British Library; Dr. Carlos Falci, Associate Professor at the School of Fine Arts at Federal University of Minas Gerais, Brazil, Visiting Fellow IAS Warwick University.

[2] Access to edit mode: http://mrl.li/mDaKJV9G; access to view mode: http://mur.al/bLZlVkD5

[3] http://soundmattersframework.wordpress.com

[4] Dr. Frederic Font, Music Technology Research Group link=http://www.mtg.upf.edu/, University Pompeu Fabra; Trond Lossius, Bergen Center for Electronic Arts http://www.bek.no/front?locale=enoct - Norway; Chris Baume, BBC R&D, http://www.bbc.co.uk/rd/blog/2015/05/spoken-words-and-their-timings, and University of Surrey; David Moffat, Centre for Digital Music (C4DM) Research Group at Queen Mary University of London

[5] http://jamoma.org

[6] http://mukurtu.org

[7] http://freesound.org

[8] http://jamoma.org/modular/

[9] From 'Poetics of Memory: invention and discovery using metadata to create cultural memories in programmable environments', by Dr. Carlos Falci

[10] The interview was conducted in Warwick University on the 18th of March, 2015.

[11] Michael Gallagher's proposal of method for annotating sound: causal listening, to determine what caused the sound; semantic listening, to determine what the sound means; reduced listening, which attends to the aesthetic characteristics of the sound; associative listening, memories and associations evoked; affective listening, how the body feels within sound; spatial listening, sensing the direction and distance of sound sources in space in relation to the listener; critical listening, listening for the wider social, economic, historical and political context of sound – which may also involve gathering additional information." By Michael Gallagher in "Audio Methods: Analysing field recordings of electronic voices in Athens and Glasgow

[12] Andrea Zarza, Sonic Time Capsule: https://ia902609.us.archive.org/16/items/SonicTimeCapsule/sonic%20time%20capsule.pdf Accessed on 20/05/15

[13] Examples of Freesound API can be seen here: http://labs.freesound.org/api_demos

References and bibliography

Celma Herrada, Oscar (2008) Music Recommendation and Discovery in the Long Tail. PhD Thesis. University Pompeu Fabra.

Falci, Carlos Henrique Rezende (2014), "Poetics of Memory: invention and discovery using metadata to create cultural memories in programmable environments" (document given by the author)

Font Frederic, Roma Gerard & Serra Xavier (2013) Freesound Technical Demo. MM'13, Ocotober 21-25, 2013, Barcelona, Spain. ACM 978-1-4503-2404-5/13/10. http://dx.doi.org/10.1145/2502081.2502245

Gallagher, Michael (2013) "Audio Methods: Analysing field recordings of electronic voices in Athens and Glasgow" (document given by the author)

King, Mari & McKernan Luke (2013), Opening up Speech Archives

Mechtley Brandon, Cook Perrey & Spanias, Andreas (2012) "Shortest Path Techniques for Annotation and Retrieval of Environmental Sounds" International Society for Music Information Retrieval.

Roma G. , Janer Jordi, Kersten Stefan, Schirosa Mattia, Herrera Perfecto, and Serra Xavier (2010) "Ecological Acoustics Perspective for Content-Based Retrieval of Environmental Sounds" Hindawi Publishing Corportaion. *EURASIP Journal on Audio, Speech, and Music Processing.* Vol 2010, Article ID 960863, 11 pages. Doi:10.1155/2010/960863