Innovative Tools for Sound Sketching Combining Vocalizations and Gestures

Olivier Houix
STMS Ircam-CNRS-UPMC,
Perception and Sound Design
1, place Igor Stravinsky
75001 Paris, France
olivier.houix@ircam.fr

Stefano Delle Monache
Iuav University of Venice
Department of Architecture and Arts
Dorsoduro 2196
30123 Venezia, Italy
sdellemonache@iuav.it

Hélène Lachambre
Genesis
Domaine du petit Arbois - BP 69
13545 Aix-en-Provence Cedex 4, France
hlachambre@optis-world.com

Frédéric Bevilacqua
STMS Ircam-CNRS-UPMC,
Sound Music Movement Interaction
1, place Igor Stravinsky
75001 Paris, France
frederic.bevilacqua@ircam.fr

Davide Rocchesso
Iuav University of Venice
Department of Architecture and Arts
Dorsoduro 2196
30123 Venezia, Italy
roc@iuav.it

Guillaume Lemaitre
STMS Ircam-CNRS-UPMC,
Perception and Sound Design
1, place Igor Stravinsky
75001 Paris, France
guillaumejlemaitre@gmail.com

ABSTRACT

Designers are used to produce a variety of physical and digital representations at different stages of the design process. These intermediary objects (IOs) do support the externalization of ideas and the mediation with the different stakeholders. In the same manner, sound designers deliver several intermediate sounds to their clients, through iteration and refinement. In fact, these preliminary sounds are sound sketches representing the intermediate steps of an evolving creation. In this paper we reflect on the method of sketching sounds through vocalizations and gestures, and how a technological support, grounded in the understanding of the design practice, can foster transparency and mediation in sound design-thinking. Three tools, under development in the scope of the EU project SkAT-VG (Sketching Audio Technologies using Vocalizations and Gestures) are introduced and discussed, based on the preliminary observations collected during a workshop involving professional sound designers.

CCS Concepts

•Human-centered computing → Interaction design process and methods; Systems and tools for interaction design; •Applied computing → Sound and music computing; •Hardware → *Sensor devices and platforms; Sound-based input / output;*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AM '16, October 04-06, 2016, Norrköping, Sweden

© 2016 ACM. ISBN 978-1-4503-4822-5/16/10…$15.00

DOI: http://dx.doi.org/10.1145/2986416.2986442

Keywords

Sound Design, Vocalization, Gesture, Interaction

1. INTRODUCTION

Designing an object, a service, or other physical or digital artifacts is a practice grounded on project, namely articulated between conception and realization parts. The project can be divided in different stages: Inspiration; Understanding the context and the problem; Ideation as an iterative process of generating and evaluating ideas, Implementation of concrete solutions [4] .

During the different stages of work, designers usually produce a large set of physical or digital representations, in the form of sketches, rough technical drawings, perspective projections, or 3D mock-ups. These intermediary objects (IO) of design [3], or boundary objects [6], represent a crucial support to the individual activities of the designer, not only as means to externalize her ideas and intentions [2], but also to support mediation between the different stakeholders, for instance the partners of different departments (marketing, graphic designers, engineers, and so forth), during the several states of a design project.

Similarly to the industrial designer, the sound designer shapes ideas or guidelines from the design brief into sound proposals, through an iterative process of refinement; for example Susini et al. [33] proposed an iterative process based on the 3-step paradigm of analysis, creation and testing.

The preliminary or intermediate sound proposals are not to be confused with the final sound design delivered to the client, as they are meant to enable shared discussion among the different stakeholders involved, sound professionals and not. Hence, these sound proposals are rather to be considered as sketches, and interpreted as intermediate steps of the evolving creative process, in the same manner as the drawings of a product indicate the progression of the creative thinking.

In this paper, we reflect on the practice of sound sketch-

ing through vocalizations and gestures. First, we discuss the concept of sound sketching and its properties in comparison with the visual domain. We discuss vocal sketching as a method, and introduce three software-based tools, miMic, MIMES and SkAT Studio, which support sound sketching by coupling vocalization and gestures. We further elaborate our reasoning on the tools, based on the outlined properties of sketches. In the conclusions, we show how the different properties and affordances of the technological tools can support the production of sound propositions that are understood as sketches. In addition, we open room for discussion around the role of the tools as means to foster the design process.

2. WHAT IS SKETCHING?

2.1 In the Visual Domain

Design is not an activity performed by a "lone creative genius" [4], but a process involving a multidisciplinary teamwork. In this context, IOs are relevant means to mediate the activity among partners with different terminology and cultures.

An IO may be described along two dimensions. The first dimension is linked to the communication between the designer and the stakeholders and/or users, and the second one is related to the appropriation of the IOs by the stakeholders or users [2, 3]. IOs may combine different properties along these two dimensions depending on the stage of the design process.

In order to facilitate appropriation, an *open* IO is prone to modifications in order to adapt the design to the context of the project. Conversely, a *closed* object may accept no modification to avoid disrupting the design process.

Intermediary objects favor the communication between the designer and the stakeholders (e.g., potential end-users). IOs are *transparent* when the designer intentions are fully conveyed and understood through and with the use of the IO. On the other hand, IOs take on a mediation role, when the stakeholder's interaction and participation suggests new or diverse design directions [2, 3].

For example, a closed and transparent object is typically based on standardized forms or methods, like a technical drawing, in order to transmit a prescription at the end of the project. An open and transparent IO is generally a representation with a shared language or system like digital tools (Computer-Aided Design), or 3D mock-up [2] that foster a collaborative work and an interdisciplinary dialogue.

Considering 2D intermediary objects, usually encountered in the early design stages, drawn sketches with pencil on paper are externalized ideas in the form of configurations of simple elements and spatial relations [35].

Sketching has been described as possibly assuming four roles [36]: i) the *thinking sketch* as a direct support to the individual design-thinking process, where the designer and the sketch are in a continuous loop of understanding while crafting; ii) the *talking sketch* as a hook to foster discussion among peers in the design process; iii) the *prescriptive sketch* which is aimed at communicating design decisions to actors not directly involved in the design process, for instance to the manufacturer; and finally iv) the *storing sketch*, allowing designers to archive their ideas and return to them later. To summarize, sketches are both a process and a product that allow a reflective conversation and that elicit interpretation and condensation of meaning [10].

Buxton [5] described the different properties of an effective sketch. Sketching is quick, timely, inexpensive and disposable. It is generally prolific, that is it contains plenty of proposals, characterized by clear vocabulary and conventions, with distinct meaningful strokes (gestures), yet within a sense of openness and freedom. Sketches provide minimal details, yet affording an appropriate degree of refinement. They suggest and explore rather than confirm or tell. It is the inherent ambiguity of sketches that opens spaces to interpretation.

Up to this point, we have been presenting ideas that refer mainly to visual sketching, but what are the roles and properties of sound sketches in comparison with more traditional visual approaches?

2.2 Sound Sketching

Schön and Wiggins [29] described the design process as a succession of "seeing, moves and seeing" steps. The designer approaches the 2D or 3D physical embodiment of her ideas in a reflective conversation, in which inspection and action alternate in steps of problem-setting and analysis, crafting, and evaluation.

Similarly, Nykänen et al. [22] analyzed the sound design process of car logotypes. They suggested a similar design process of iterative hearing and doing. However, whereas one may argue that the visual and the sound design processes are essentially superimposable, the authors pointed out that in the current sonic practices designers are not really acquainted with the production of sound sketches. Instead, the sound materials they propose, being them thinking, talking, prescriptions or records, are more likely to be closer to the status of final proposals or prototypes. The relative novelty of sound design in the industry makes sketching, which is a well-established process in the visual design domain, a rather abstract concept, still.

Schön and Wiggins claimed that "designers work in medium" [29, p. 154] and the sketch is typically a medium that supports ideation, or a medium through which designers hear an evolving product of their work [22]. Steenson and Rodger [32] consider sound not only as a "material", but also as a sketch of possible affordances, bringing sound into use.

However, concerning sound design, is sound really a medium to support sketching, that is facilitating the understanding of 2D or 3D intermediary objects?

In fact, one fundamental difference between visual and sound sketches is that drawing deals with spatial relations, whereas sonic configurations essentially emerge in time. Secondly, whereas there is a clear distance between the visual sketch (i.e., the drawing) and the final design in terms of presentation (e.g., from the drawing of an artifact, to its photorealistic rendering, to the actual physical realization), a similar distance is not always as apparent in current sound sketching and design. Finally, shared conventions on sound are not often found between the different stakeholders, who may not even agree on a common lexicon about sound [7].

Delle Monache et al. [10] outlined the basic properties of the sound sketch, from the cumulative experience of several workshops on sonic interaction design. Similarly to the visual domain, in which apparently complex illustrations can be drawn by juxtaposing and overlaying a limited set of basic shapes, sonic representations are effectively articulated

by means of a composition of similar basic sonic elements, elementary physical interactions [18, 1] and morphological profiles. These basic elements become words of the vocabulary available to the sound designer.

Apparently, the aesthetics of sound represents the primary concern preventing the sound designers to fully exploit sketching in their practice. However, effective and meaningful sound sketching requires fluency and expertise, methods and tools affording the manipulation of the representation without sparing attention on the production process. In other words, the aesthetic quality of the sound sketch is a partially ill-posed issue, as the beauty or clarity of the sound rendering is not an indicator of the step in the process. Dexterity in sound sketching does not equal to proficiency of rendering.

Based on these observations, we propose a characterization of what effective sound sketches should be. Effective sound sketches should be provisional, incomplete, yet meaningful representations with an aesthetics of its own, that can be easily transformed, modified, stored, recalled and tracked. As such, they do not exist per se, instead they are part of the wider workflow of creativity and design, which implies the cooperation with peers and stakeholders.

2.3 Sound Sketching Tools

Sound design practice is a highly conceptual and ill-defined activity, which is further undermined by the lack of a shared vocabulary on sound, among practitioners and stakeholders. Professionals leverage the technical jargon to specify their verbal concepts, which is not accessible to lay-persons [24].

Communication tools, such as a deck of cards or a sound charter, have been proposed to make design conversations about sound more effective and reliable [7]. These cards are closed IOs, nevertheless acting as mediators as they foster communication and co-design. On the other hand, humans are pretty good at describing sounds by means of vocal imitations, and such ability may be exploited in the early stages of the sound design process [20].

A decade of research in sonic interaction design has been developing a plethora of approaches to sound sketching and design, aimed at fostering creativity and understanding, and yet stressing the instrumental role of technology. For example, Hug and Kemper [16] developed a framework to foster creativity in the development of new interactive commodities. Their framework exploits Foley-based[1] techniques to design sound in realtime, and Wizard-of-Oz approaches to deploy the sound sketches in the demonstration of the interactive, electroacoustic mock-ups. The pedagogical value of this framework stresses the role of the resulting mock-ups as open IOs easily available to modifications within the design session.

PSST! [17] is a tool that uses tangible objects, like in the Reactable[2], to shape sound by combining granular synthesis and sampling techniques. However, this tool is relatively closed, mostly suitable to produce sound alternatives rather than sound ideas from scratch.

Increasing research efforts are being directed toward the exploitation of vocalizations and gestures as basic conceptual tool for sound sketching [11, 25]. Ekman and Rinott [11] proposed the use of vocalizations as thinking and communication tool to support the fast and rough prototyping of sound ideas in the early stage of the design process. They proposed vocal sketching as a cooperative tool to enact teamwork in sonic interaction design. One drawback is the intrinsic limit of the human vocal apparatus to the production of a limited set of myoelastic and turbulent sounds.

VOGST - Voice Gesture Sketching Tool[3] is an experimental, hand-held device which attempts to overcome the ephemerality of vocal sketches, by adding the possibility to record the vocalization and the gesture, and playback and manipulate them creatively. Vocalizations and gestures become mediating IOs, through the technology, and open the design space to possible modifications, even through collaborative interaction.

Tahiroğlu et al. [34] developed a graspable mobile device that captures touch, squeeze and motion. Vocalizations and gestures were exploited in the conceptual stage of two design tasks, the first concerning simple communication like sending a message, and the second related to expressive musical interaction. They developed a gestural vocabulary, based on basic motions such as shaking, pointing, circular movements, and coupled it with vocal sketches of real world and abstract sounds. For example pointing out a direction was associated with discrete pitched sounds.

Vocalizations and gestures have been also exploited as a way to navigate databases or to control sound synthesis in realtime. SynthAssist [8] is a sound synthesis production tool based on the audio information retrieval of example recordings, such as vocal imitations. The user provides a probe vocalization, and compares and rates the system suggestions, in order to navigate a database of pre-synthesized sounds. The iteration of the rating process allows to reach proximity to the desired sound. Esling and Agon [12] proposed also a query by vocal imitation system which allows users to search sound across a sound database through their imitation.

Another example is the VocaListener [21], a singing synthesis system that automatically estimates parameters for singing synthesis from the user's singing voice.

Finally vocalizations and gestures are the main tools investigated in the FP7 FET-Open EU project SkAT-VG, aimed at expanding the knowledge on the perception and non-verbal communication of sonic concepts between humans, and at developing sketching tools for sound design [19, 25]. The basic knowledge on the perception and production of vocal imitations and gestures is distilled in applications and tools to support the conceptual stage of the sound design process. In the next section, we present and analyze three demonstration tools that have been developed during the project.

3. VOICE AND GESTURE AS A SKETCHING TOOL

3.1 The Practice of Vocal Sketching

A framework on the use of voice and gestures as means to specify sound ideas in the early stage of the design process has been devised in several workshop experiences. The workshop framework combines introductory exercises to vocalization techniques, warming up, training in the form of a guessing game on vocal imitations, and several fictional design tasks, in a gamification environment aimed at stressing

[1] https://en.wikipedia.org/wiki/Foley_(filmmaking).

[2] https://en.wikipedia.org/wiki/Reactable.

[3] http://blogs.iad.zhdk.ch/vogst/.

the cooperation value, the expressiveness and the immediacy of sketching in the sound design domain. For example, the value of cooperative teamwork is stressed in a design task involving video-prototyping of motor sounds, under pressing time constraints [10].

The effectiveness of the design exercises is supported by the evidence collected in a set of behavioral experiments in the lab. In one experiment, subjects were asked to imitate a set of referent sounds (combinations of pure tones and narrowband noises), with the goal of understanding to what extent basic audio features (pitch, tempo, sharpness and onset) are correctly reproduced for the recognition of the referent sound. The analysis of the vocal imitations indicate that pitch and tempo are reproduced accurately with the absolute values of the features, the vocal imitations of sharpness showed a transposition of the feature into the participants registers, and finally vocal imitations of onsets categorized the continuum of onset values into two discrete morphological profiles [19]. Hence, when imitating referent sounds, humans tend to emphasize their salient features within the constraints of human vocal production.

Taken together, vocal sketches represent fluent, minimal means for a faithful representation of a given sound. The vocal sketch enacts *thinking* and *talking*, although its inherent ephemerality limits the possibility to trace and store the process, and restricts the space of meaningful transformations [31]. In other words, the vocal sketch shows a dual nature, as inherently transparent means embodying the intention of the sound designer, yet as an external representation hardly changeable by other partners (i.e closed IOs). In the next subsection we present some technological approaches that extend the potential of vocal sketching, by exploiting vocal and gestural imitations to drive synthetic sound models. We elaborate on these tools, by stressing their potential as IOs for communication and appropriation, and their thinking, talking, prescriptive, and storing characteristics.

3.2 miMic

miMic is an eyes-free interface for vocal and gestural sketching. The system[4] is composed of a microphone augmented with embedded inertial sensors and two buttons for contextual functions (i.e., train, set, play modes). It makes use of a software which handles the classification of the vocal imitations as instances of a palette of physics-based sound models (implemented in Max by Cycling 74[5]), that can be set and played [26, 1] .

miMic, shown in figure 1, is the pencil in the hand of the sound designer, who can vocalize a sound idea and convert it in a mixture of synthetic sound models (set mode). The resulting configuration of synthetic sound primitives defines a sound design space that can be further played and explored creatively via vocal and gestural control[6] (play mode). In addition, the configuration of sound models can be further refined in terms of behavior and timbre, by tweaking on GUI either the vocal features to models parameters maps or the non-interactive parameters.

The eyes-free operation in train, set and play modes is aimed at fostering fluency and immediacy, by stressing the full integration of the technological support to vocal sketching, within the creation workflow. The resulting sound is a rough and basic, yet meaningful representation, as the configuration is composed of a mixture of perceptually-relevant sound primitives (e.g., impacts, frictions, liquids, gasses, machines) [1]. Moreover, the fine-tuning of the configurations on GUI allows to refine to a certain extent the sound quality, for instance when playing the Wizard-Of-Oz in sonic interaction design teamwork. In this respect, the miMic system fosters transparency and mediation, and facilitates the quick production of IOs that can be layered, stored, and recalled in the form of configurations of sound models or audio files.

Figure 1: miMic.

3.3 MIMES

A further approach to sound sketching is based on the complementary use of vocalizations and gestures. In the MIMES system, vocal sound ideas are automatically analyzed and acquired as time/morphology profiles of acoustic features (loudness, spectral centroid, spectral skewness, spectral spread, noisiness and zerocrossing) [23]. The resulting lists of features are used as input parameters to drive corpus-based sound synthesis [30], similarly to real-time audio mosaicing [28]. More precisely, these voice descriptors are used to select small sound samples from a database of sounds (corpora) with similar descriptors (using the *MuBu & friends* library in Max [27]). This process generates a synthetic sketch with similar audio features (which can be selected).

The use of different audio databases allows the designer to explore creatively different sound qualities, for example by associating either abstract or concrete sounds to the initial vocal sketch. Thus, the designer can test such different sound rendering and eventually mix them. It is also possible to manipulate the weight of each sound descriptors (i.e., stressing a given descriptor by tweaking its weight).

The designer can also use gestures/actions to further transform the sound sketch. This is concretely produced by manipulating a tangible object [15], as the one shown in Fig-

[4] http://buildinprogress.media.mit.edu/projects/2385/steps

[5] https://cycling74.com

[6] https://vimeo.com/142351022

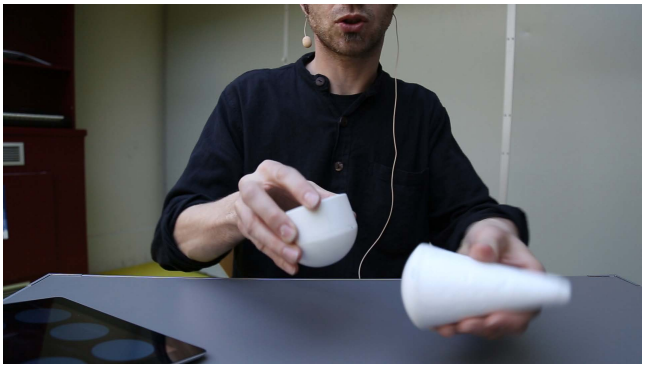

Figure 2: MIMES.

ure 2. Sensors are embedded in the tangible object, including a 9 degree-of-freedom inertial motion unit (accelerometers, gyroscope, magnetometers), a FSR (Force Sensitive Resistor) and piezo sensors. The sensors data are analysed in real-time to recognise different "playing techniques", which are associated to different sound transformations[7]. Moreover, the "mapping by demonstration" technique [14, 13] allows for adding new user-defined gestures to sound effects.

This tool could be fluently integrated during the sound design process in exploring different sound textures or combinations of textures, as a *thinking* sketch. An interesting possibility for the sound designer is the use of his own sounds, and the inclusion of large sets of sounds. The sound result could either be rough or relatively complex depending on the sound corpora used and the possible exploration of the sound variations within a corpus. Moreover, the designer can produce simple sonic forms or craft complex time morphologies (with the voice and with gesture) that can be stored and readily exploited in interaction design contexts. The embodied character of MIMES makes it suitable not only for self-reflection, but also as *talking* or *prescriptive* tool for interaction in a design team. Possibly, it can be used together with the customer, in co-design sessions, thus acting as a mediating and open IO.

3.4 SkAT Studio

SkAT Studio is a Cycling 74 Max modular framework, intended to work as a host of diverse technologies for vocal sketching, like miMic and Mimes. The general workflow is composed of 5 stages, as described in Figure 3: i) input, that is vocal and gestural signals; ii) analysis, for audio and gestural features extraction; iii) mapping, that is the transformation of the extracted features in control values for sound production; iv) sound production, synthesis and processing; v) output, namely sound rendering and recording.

In the main GUI, shown in figure Fig 4, it is possible to load a collection of modules, according to the general workflow. Each group, representing a stage of the workflow, can host multiple modules simultaneously (i.e., multiple feature extractors, control maps, and sound synthesizers), and signal and control data routing throughout the SkAT Studio framework can be programmed by means of patchbays. miMic sound synthesizers [1] have already been included in SkAT Studio framework. In general, modules (analysis, mapping, sound synthesis and processing) can be easily implemented as MAX patches, according to a template that is provided with the software release.

A configuration of modules in groups defines a "control scenario", that can be stored locally and recalled as a global preset. Control scenarios can be loaded manually from disk, or through vocal imitations by exploiting the classification put on top of SkAT Studio, in a manner similar to miMic architecture. The number of loadable modules (i.e., sound models) per vocal imitation input has been limited to the best three results. In addition, the relative weight of the three sound models can be further tweaked by hand.

Compared to miMic and MIMES, SkAT Studio represents a sort of desktop tool for sound sketching, providing a rather detailed control over the working chain, and getting closer to the operation of traditional DAWs In fact, the modules in the groups can be potentially replaced by plugins, according to the standard formats. SkAT Studio GUI allows to play the configurations through voice and gestures, and optionally by on-screen interaction with the mouse. The interactive parameters are provided with a freeze function to bypass the vocal and/or gestural control, a simple functionality to facilitate the creative production of raw sound materials. Moreover, the design process is tracked in the form of history of actions performed during the session, that can be navigated back and forth. However, one drawback, possibly affecting mediation, is that the general modularity requires the programming of control scenarios, before-hand. This limitation is reduced with the progressive development and collection of control scenarios.

4. PERSPECTIVES

SkAT Studio and MIMES were recently proposed and used as main sketching tools during the "48h Sound Design" Workshop at Château La Coste art park and vineyard, in Le Puy Ste Réparade, France[8]. Five professional sound designers were invited to work each on one of the site-specific art pieces located in the park, and design an accompanying sound signature for the chosen art installation, in 48 hours. The sound designers received a half-day crash training on vocal sketching and the tools, in the day before the actual start of 48h of Sound Design. They were required to sketch the sound ideas in the first day, with the exclusive use of the tools provided (only MIMES and SkAT Studio were provided). In the second day, each sound designer could use his own tools to refine the sound sketch selected with a stakeholder expert of the art piece.

Since the analysis and evaluation of the workshop is out of the scope of this paper, here we report some preliminary observations on the sketching tools, emerged during the debrief held with the sound designers, the day after the workshop.

The sound designers stressed the complementary nature of the two tools. These were generally experimented in a workflow of quick creation of rough sound materials in SkAT Studio, which were lately imported as sound corpora for further textural processing and refinement in MIMES.

In their current release, the tools allowed to export sound ideas in the form of sound files on disk. This was perceived as a major limitation affecting the ease of their workflow, given the potential of the procedural audio approach in SkAT Studio and MIMES. As seen from a different angle,

[7] Video presenting a proof-of-concept: https://vimeo.com/125367426

[8] https://vimeo.com/169521601.

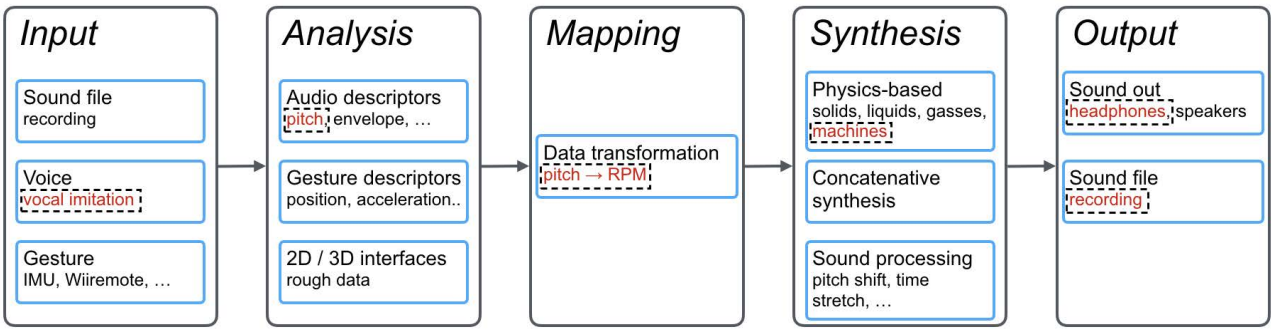

Figure 3: Skat Studio workflow. In dashed boxes, a basic configuration of voice-driven engine sound synthesis.

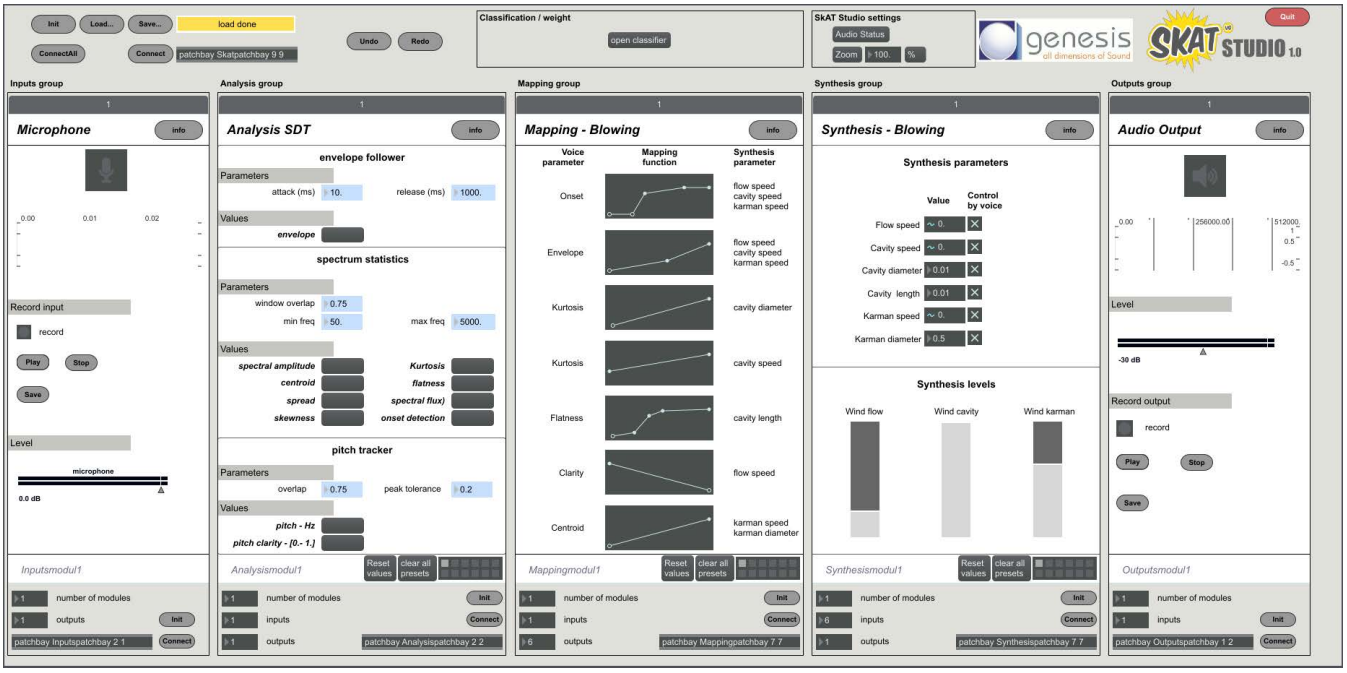

Figure 4: Skat Studio GUI, control scenario for the voice-driven blowing sound synthesis.

one can argue that the inherent malleability of audio procedures and models makes them a rather effective *talking* means from which deriving new and improved early sound ideas (i.e., lateral and vertical transformations). Conversely, sound files may represent better means to embody *prescriptive* and *storing* elements. In this respect, the sound designers largely exploited the available presets functions (i.e., store and recall), that however could not be exported nor integrated in their DAWs.

These vocal imitation-to-sound (model) converters proved to be an effective means to ease non-verbal thinking, yet the actual palette of sound models was found to be too bounded to realistic behaviors[9]. Similarly, it was found as difficult to tailor SkAT Studio's behavior in terms of grain modeling. These observations go back to the long-standing issue of control strategies, whether user-centered or general purpose, pre-defined or malleable, and the time-consuming effort required for their tuning. Since one of the objectives is to provide tools which would not divert attention from the primary activity of sound sketching, we set a software environment with predefined control maps, exploit them as creative constraints, and let the designers learn and adapt to the expression of the tools.

Finally, as a general remark, the technological support to vocal sketching was received positively, especially in view of having a personal, eyes-free, interactive audio sketchbook. On the contrary, the designers showed their skepticism towards the feasibility of participatory vocal sketching processes involving clients and stakeholders. They were especially concerned by the inherent low quality of the sound sketches, yet their reluctance rather reflected the mastering of the methods.

[9] However, the proof-of-concept sound installation "S'í fosse suono" (If I were sound) is the countercheck: several vocal sketches of audio self-portraits have been re-synthesized by exploiting physics-based or concatenative synthesis. http://www.skatvg.eu/SIFosse/

5. CONCLUSION

Intermediary sonic representations, both physical and digital, are essential means to embody concepts in design materials (*thinking* sketch), to foster communication and cooperation among peers (*talking* sketch) and with the other stakeholders (*prescriptive* sketch).

The lack of a shared terminology and culture on sound in the design domain, prompted us to develop new methods and practices for sound sketching grounded on vocalizations and gestures.

Vocalizations are easy to produce, and represent preliminary realizations in the process of sound creation, Vocalizations are *transparent* means available to the sound designer to express her intentions. However, the major drawback is the effort required by the other partners to access and effectively manipulate and modify this kind of representation.

Although good imitations of sounds are often achievable with practice, a bad imitation could be counterproductive in design negotiations. However, the creative potential of bad imitations in sound design and composition has been previously highlighted [9], and supports the use of vocalizations as means to foster ideas generation and brainstorming.

In miMic, voice and gestures are coupled and exploited to access a rather immediate control over complex sound synthesis models. The sound designer can easily produce rough sound sketches by modulating and controlling the sound behavior dynamically with voice and gesture, in order to contextualize sonic interactions within the use case at hand. MIMES allows to explore the sonic space of combinations of textures, through vocalizations and gestures. The tangible interaction in MIMES stresses the cooperative dimension of vocal sketching, as expressive sound sketches can be refined collaboratively. A common framework, SkAT Studio has been developed to integrate both approaches to sound creation, and provide an overall environment which affords sound sketching at various levels of detail.

The three tools were received positively by sound designers, yet their design should promote a rather effective integration in the workflow. Based on these early observations, we plan to assess the software integration, and their emerging properties systematically, in sound design tasks with expert practitioners.

6. ACKNOWLEDGMENTS

The work described in this paper is part of the project SkAT-VG, which received the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission under FET-Open grant number: 618067.

7. REFERENCES

[1] S. Baldan, S. Delle Monache, D. Mauro, and D. Rocchesso. Automatic system for the generation of sound sketches. Technical report, SkAT-VG Project, December 2015. Deliverable D6.6.1.

[2] J.-F. Bassereau, R. C. Pello, J. Faucheu, and D. Delafosse. Les objets intermédiaires de conception/design, instruments d'une recherche par le design. *Sciences du Design*, 2:48–63, 2015.

[3] J.-F. Boujut and E. Blanco. Intermediary objects as a means to foster co-operation in engineering design. *Computer Supported Cooperative Work (CSCW)*, 12(2):205–219, 2003.

[4] T. Brown. Design thinking. *Harvard business review*, 86(6):84, 2008.

[5] B. Buxton. *Sketching User Experiences. Getting the Design Right and the Right Design*. Elsevier / Morgan Kaufmann, Boston, MA, USA„ 2007.

[6] P. R. Carlile. A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization science*, 13(4):442–455, 2002.

[7] M. Carron, F. Dubois, N. Misdariis, C. Talotte, and P. Susini. Designing sound identity: Providing new communication tools for building brands "corporate sound". In *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*, AM '14, pages 15:1–15:8, New York, NY, USA, 2014. ACM.

[8] M. Cartwright and B. Pardo. SynthAssist: Querying an Audio Synthesizer by Vocal Imitation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 363–366, London, United Kingdom, June 2014. Goldsmiths, University of London.

[9] A. Cera. Loops, games and playful things. *Contemporary Music Review*, 32(1):29–39, 2013.

[10] S. Delle Monache, D. Rocchesso, S. Baldan, and D. Mauro. Growing the practice of vocal sketching. In K. Vogt, A. Andreopoulou, and V. Goudarzi, editors, *Proc. of the 21th International Conference on Auditory Display (ICAD-2015)*, pages 59–65, Graz, Styria, Austria, 2015.

[11] I. Ekman and M. Rinott. Using vocal sketching for designing sonic interactions. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, pages 123–131. ACM, 2010.

[12] P. Esling and C. Agon. Multiobjective time series matching for audio classification and retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(10):2057–2072, Oct 2013.

[13] J. Françoise, N. Schnell, and F. Bevilacqua. A Multimodal Probabilistic Model for Gesture–based Control of Sound Synthesis. In *Proceedings of the 21st ACM international conference on Multimedia (MM'13)*, pages 705–708, Barcelona, Spain, 2013.

[14] J. Françoise. *Motion-sound Mapping By Demonstration*. Theses, Université Pierre et Marie Curie - Paris VI, Mar. 2015.

[15] O. Houix, F. Bevilacqua, N. Misdariis, P. Susini, E. Flety, J. Françoise, and J. Groboz. Objects with Multiple Sonic Affordances to Explore Gestural Interactions. In *xCoAx conference, Computation, Communication, Aesthetics & X*, pages 296–303, Glasgow, UK, June 2015. Alison Clifford, Miguel Carvalhais & Mario Verdicchio.

[16] D. Hug and M. Kemper. From foley to function: A pedagogical approach to sound design for novel interactions. *Journal of Sonic Studies*, 6(1), 2014.

[17] R. J. Jansen, E. Özcan, and R. van Egmond. Psst! product sound sketching tool. *Journal of the Audio Engineering Society*, 59(6):396–403, 2011.

[18] G. Lemaitre, O. Houix, N. Misdariis, and P. Susini. Listener expertise and sound identification influence

the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, 16(1):16, 2010.

[19] G. Lemaitre, A. Jabbari, N. Misdariis, O. Houix, and P. Susini. Vocal imitations of basic auditory features. *The Journal of the Acoustical Society of America*, 139(1):290–300, 2016.

[20] G. Lemaitre and D. Rocchesso. On the effectiveness of vocal imitations and verbal descriptions of sounds. *The Journal of the Acoustical Society of America*, 135(2):862–873, 2014.

[21] T. Nakano and M. Goto. Vocalistener: A singing-to-singing synthesis system based on iterative parameter estimation. *Proc. Conference on Sound and Music Computing*, pages 343–348, 2009.

[22] A. Nykänen, J. Wingstedt, J. Sundhage, and P. Mohlin. Sketching sounds – Kinds of listening and their functions in designing. *Design Studies*, 39:19–47, July 2015.

[23] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams. The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5):2902–2916, 2011.

[24] T. Porcello. Speaking of sound language and the professionalization of sound-recording engineers. *Social Studies of Science*, 34(5):733–758, 2004.

[25] D. Rocchesso, G. Lemaitre, P. Susini, S. Ternström, and P. Boussard. Sketching sound with voice and gesture. *interactions*, 22(1):38–41, 2015.

[26] D. Rocchesso, D. A. Mauro, and S. Delle Monache. mimic: The microphone as a pencil. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '16, pages 357–364, New York, NY, USA, 2016. ACM.

[27] N. Schnell, A. Röbel, D. Schwarz, G. Peeters, and R. Borghesi. MuBu and Friends – Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP. In *Proceedings of the International Computer Music Conference (ICMC 2009)*, pages 423–426, Montreal, Canada, 2009.

[28] N. Schnell, M. A. Suarez-Cifuentes, and J.-P. Lambert. First Steps in Relaxed Real-Time Typo-Morphological Audio Analysis/Synthesis. In *Proceedings of Sound and Music Computing Conference (ICMC 2010)*, pages 1–1, Barcelone, Spain, 2010.

[29] D. A. Schön and G. Wiggins. Kinds of seeing and their functions in designing. *Design studies*, 13(2):135–156, 1992.

[30] D. Schwarz. Corpus-based concatenative synthesis. *IEEE Signal Processing Magazine*, 24(2):92–104, 2007.

[31] H. Scurto, G. Lemaitre, J. Françoise, F. Voisin, F. Bevilacqua, and P. Susini. Combining gestures and vocalizations to imitate sounds. *The Journal of the Acoustical Society of America*, 138(3):1780–1780, 2015.

[32] J. Steenson, M. Rodger, and W. Matthew. Bringing sounds into use: Thinking of sounds as materials and a sketch of auditory affordances. *The Open Psychology Journal*, 8(1), 2015.

[33] P. Susini, O. Houix, and N. Misdariis. Sound design: an applied, experimental framework to study the perception of everyday sounds. *The New Soundtrack journal - Special Issue: Perspectives on Sound Design*, 4:103–121, 2014.

[34] K. Tahiroğlu and T. Ahmaniemi. Vocal sketching: a prototype tool for designing multimodal interaction. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, page 42. ACM, 2010.

[35] B. Tversky. What do sketches say about thinking. In *2002 AAAI Spring Symposium, Sketch Understanding Workshop, Stanford University, AAAI Technical Report SS-02-08*, pages 148–151, 2002.

[36] R. Van der Lugt. How sketching can affect the idea generation process in design group meetings. *Design studies*, 26(2):101–122, 2005.