SKETCHING SONIC INTERACTIONS BY IMITATION-DRIVEN SOUND SYNTHESIS

Stefano Baldan, Stefano Delle Monache, Davide Rocchesso, Hélène Lachambre
Iuav University of Venice, Iuav University of Venice, Iuav University of Venice, Genesis Aix-en-Provence
stefanobaldan@iuav.it, sdellemonache@iuav.it,
roc@iuav.it, helene.lachambre@genesis.fr

ABSTRACT

Sketching is at the core of every design activity. In visual design, pencil and paper are the preferred tools to produce sketches for their simplicity and immediacy. Analogue tools for sonic sketching do not exist yet, although voice and gesture are embodied abilities commonly exploited to communicate sound concepts. The EU project SkAT-VG aims to support vocal sketching with computer-aided technologies that can be easily accessed, understood and controlled through vocal and gestural imitations. This *imitation-driven* sound synthesis approach is meant to overcome the ephemerality and timbral limitations of human voice and gesture, allowing to produce more refined sonic sketches and to think about sound in a more designerly way. This paper presents two main outcomes of the project: The Sound Design Toolkit, a palette of basic sound synthesis models grounded on ecological perception and physical description of sound-producing phenomena, and SkAT-Studio, a visual framework based on sound design workflows organized in stages of input, analysis, mapping, synthesis, and output. The integration of these two software packages provides an environment in which sound designers can go from concepts, through exploration and mocking-up, to prototyping in sonic interaction design, taking advantage of all the possibilities offered by vocal and gestural imitations in every step of the process.

1. INTRODUCTION

Sonic Interaction Design (SID) emerged in the last few years as a new area of design science, to overcome the lack of proper design attitude and process in the exploration of innovative uses of sound for interactive products, systems and environments [1]. Its research path has been moving from the understanding of sound perception, to the definition of sound modeling approaches for design, towards a progressive, deeper understanding of how sound designers think, how they learn to think in a designerly way, and how they develop their skills and knowledge [2, 3]. The discipline proposes a systematic approach for designing acoustic interactive behaviors by means of an iterative yet linear process, made of fixed and sequential steps which emphasize the importance of the conceptual phase, the fundamental value of the expressive qualities of sound in terms of character and identity, and the holistic view of sound creation in relation to the overall design of an artefact [4].

Investigation of the early stages of the sound design process is one of the most recent and promising research tracks in this context. Like in every other design activity, *sketching* is at the core of the initial conceptual phase. Sketches are quick, disposable and incomplete representations used to embody reasoning, communicate concepts, explore divergent ideas and eventually address the design process. In visual design, pencil and paper are still the most effective sketching tools, despite all technological advances. From architectural plans to page layouts, from paper models to graphical user interface mock-ups, drawings are extensively used throughout the design process to inform and support the progressive refinement of design ideas towards the final product [5].

In the aural domain, where a direct counterpart of pencil and paper is not available yet, a promising alternative is represented by *vocal sketching*. The practice exploits the human ability in the production of non-verbal utterances and gestures to imitate the main features of a given referent sound [6]. The human voice is extremely effective in conveying rhythmic information, whereas gestures are especially used to depict the textural aspects of a sound, and concurrent streams of sound events can be communicated by splitting them between gestures and voice [7]. Despite being embodied tools, immediately available to everyone [8] and increasingly popular in education and research [9, 10], the use of voice and gesture for sonic sketching is hardly spreading among sound practitioners, especially because of the inherent ephemerality of vocal/gestural representations and because of the limited timbral palette of the human voice.

A set of interviews with eight professional sound designers was conducted by the authors, to better understand the role of sketching in sound creation practices: The conceptual phase is mostly based on browsing sound banks and/or verbally describing concepts through a lists of keywords, while sonic sketching is still a neglected practice. Pressing time constraints and the lack of a shared language between designers and clients severely affect the search quality in the conceptual phase, resulting in conservative approaches and presentation of advanced design proposals even at the

very beginning of the process. When it is used, voice mostly serves as raw material for further sound processing and rarely as real-time control, while the use of gesture is limited to the operation of knobs and faders in musical interfaces. Finally, there is a pressing and unsatisfied demand for tools which are immediate to use, providing direct accessibility to sound production and design and facilitating the time consuming activity of finding a sensible mapping between control features and synthesis parameters.

The EU project SkAT-VG [1] (Sketching Audio Technologies using Vocalization and Gesture) aims at providing sound designers with a paper-and-pencil equivalent to seamlessly support the design process from the conceptual stage to prototyping. The goal is pursued through the development of computer-aided tools, using vocal and gestural imitations as input signals to appropriately select and control configurations of sound synthesis models according to the context of use [11, 12]. These tools aim at expanding the timbral possibilities of human sound production, while retaining the immediacy and intuitiveness of vocal articulation.

The use of voice to control the production of synthesized sound has already well established foundations in the musical domain. In his PhD thesis, Janer extracts audio descriptors from singing voice for the real-time control of pitch, volume and other timbral features in physical models of actual musical instruments such as bass, saxophone and violin [13]. Fasciani proposes an interface that allows to dynamically modify the synthesis timbre of arbitrary sound generators using dynamics in the vocal sound, exploiting machine learning techniques to perform the mapping between vocal audio descriptors and synthesis parameters [14]. Analysis of gestural features and their mapping for the control of digital musical instruments is also a widely explored domain [15].

These concepts can be translated from the context of musical performance to the field of Sonic Interaction Design. Our interest is focused on vocal and gestural production which is neither organized according to musical criteria nor in verbal form, and on sound synthesis techniques for the reproduction of everyday sounds and noises rather than digital musical instruments. Such a radically different context requires novel strategies in terms of analysis, mapping and synthesis. From now on, we will refer to our approach as *imitation-driven* sound synthesis, to differentiate it from previous related work focused on musical production.

The SkAT-VG project produced at least two relevant outcomes: the *Sound Design Toolkit* (SDT), a collection of sound synthesis algorithms grounded on ecological perception and physical description of sound-producing phenomena, and *SkAT-Studio*, a framework based on sound design workflows organized in stages of input, analysis, mapping, synthesis and output. Taken together, SDT and SkAT-Studio offer an integrated environment to go from the sonic sketch to the prototype: The input stage of SkAT-Studio accepts vocal and gestural signals, which are fed to the analysis stage to extract their salient features. This higher-level description of the input is then used by the mapping stage to control the synthesis stage, which embeds SDT modules and other sound synthesis engines.

The rest of the paper is organized as follows: The Sound Design Toolkit and its software architecture are described in Section 2; SkAT-Studio is covered in detail in Section 3; Section 4 explains how the two software packages can be integrated to achieve imitation-driven synthesis; Finally, conclusions and possible future work are exposed in Section 5.

2. THE SOUND DESIGN TOOLKIT

The Sound Design Toolkit is a collection of physically informed models for interactive sound synthesis, arranged in externals and patches for the *Cycling '74 Max* [2] visual programming environment. It can be considered as a virtual Foley box of sound synthesis algorithms, each representing a specific sound-producing event.

2.1 Conceptual framework

The development legacy of the SDT [2] dates back to the foundational research on the possibilities of interaction mediated by sound, and the importance of dynamic sound models in interfaces [16, 17]. Perceptual relevance has been a key concern in the selection and veridical reproduction of the acoustic phenomena simulated by the available sound models.

In his foundational work on the ecological approach to auditory event perception, Gaver proposed an intuitive hierarchical taxonomy of everyday sounds, based on the specific properties and temporal evolution of interacting materials [18]. In his taxonomy, the whole world of everyday sounds was described in terms of solids, liquids, gases interactions, their temporally-patterned evolution, and possible compounds. For example, the sound of writing was described by a compound deformation of impacts and patterned scraping events. Similarly, the sound of a motorboat was hypothesized as a high-level combination of gases, liquids, and solids interactions.

Originally based on Gaver's work [19], the SDT taxonomy of everyday sounds has been continuously revised, updated and extended over the years, to couple the sophistication of physically informed sound synthesis with the state of the art on the perception and categorization of environmental sounds [20, 21].

The design rationale behind the organization of the provided synthesis models is to encompass a mixture of sound categories, covering the major applications of sound design that are relevant for listeners, as shown in Figure 1. Sound models are grouped according to a criterion of causal similarity (i.e., vibrating solids, liquids, gasses, and machines) and arranged in a bottom-up hierarchy. The first level presents the basic algorithms with the corresponding Max externals, suitable for the generation of a large family of simple sound events. The second level highlights the basic processes and machines (with the corresponding Max externals), that can be either straightly derived from the temporal patterning of the low-level models or that would

[1] www.skatvg.eu.

[2] http://www.cycling74.com

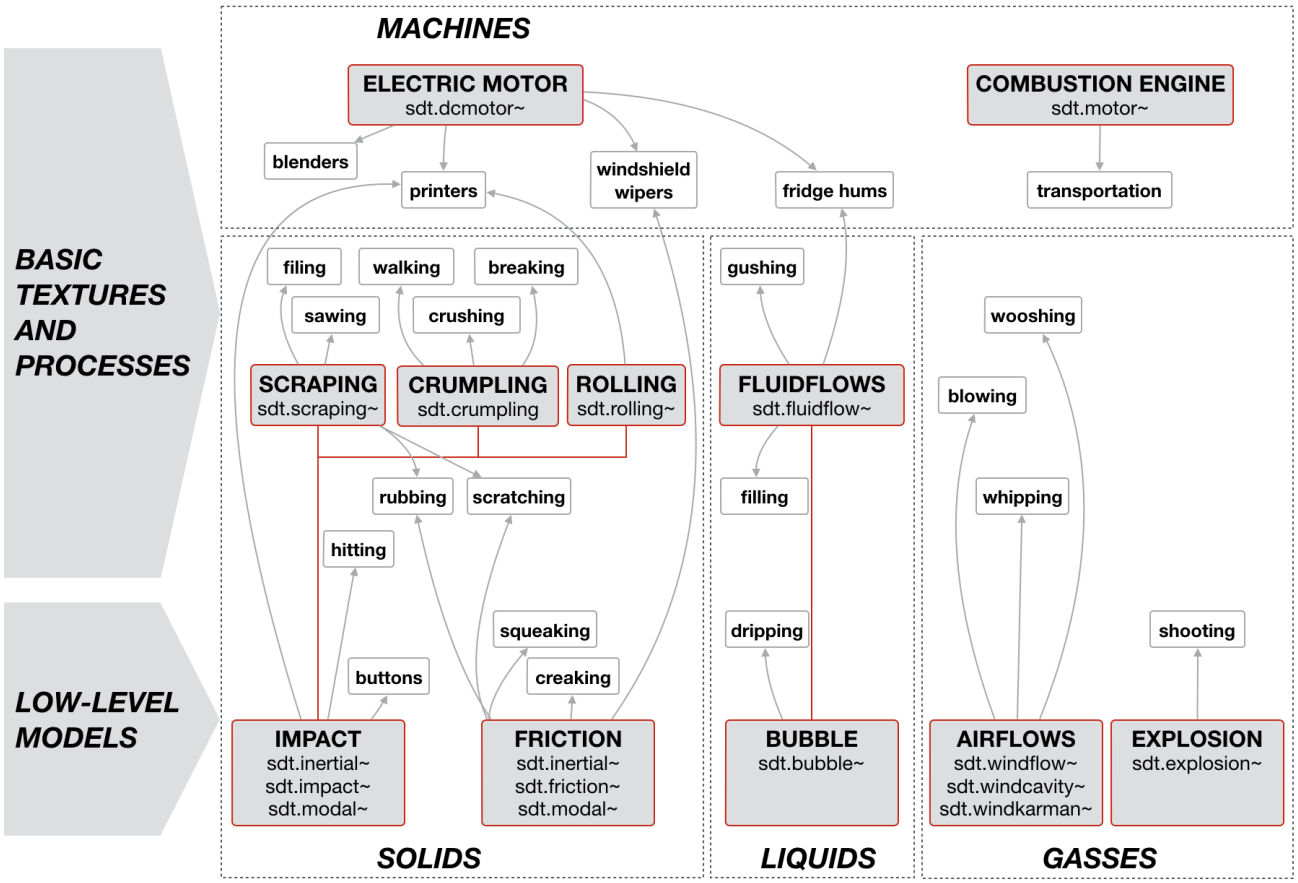

Figure 1. The SDT taxonomy of sound models. The bottom-up hierarchy represents the dependencies between low-level models and temporally-patterned textures and processes, for the four classes of sounds, solids, liquids, gasses, and machines.

be too cumbersome to develop as a Max chain of separate basic events.

In addition, the blue arrows set a direct connection between the sonic space of each model and the space of *timbral families*. Timbral families emerged from an extensive set of experiments on sound perception, as a higher-level classification of referent sounds that have been identified as cognitively stable in listeners' representations [22]. As seen from the SDT taxonomy viewpoint, a timbral family is defined as a peculiar parametrization of one or more sound synthesis models, which is unambiguously discriminated in terms of interaction, temporal and timbral properties.

2.2 Sound synthesis

The Sound Design Toolkit adopts a physically informed procedural approach to sound synthesis. In procedural audio, sound is synthesized from a computed description of the sound producing event, as opposed to sample-based techniques where sounds are prerecorded in a wavetable and then played back, manipulated and mixed together to obtain the desired timbral result [23]. Coherently with the conceptual foundation of the SDT, these computed descriptions are informed by the physics laws underlying the mechanical excitation and vibration involved in the sound events to reproduce.

The adoption of a simplified physics-based approach to sound modeling met the ecological and embodied instances emerging in computer-human interaction and design [24], thus grounding the development in design thinking and research [19]. Physically informed sound synthesis offers efficient, expressive and intuitive means to control and explore wide timbral spaces with a limited number of models, emphasizing the role of sound as a behavior, a process rather than a product. If it holds true that sound-producing events convey meaningful information about the underlying mechanical process, then manipulating their physical parameters should result in perceptually-relevant timbral modifications of the corresponding virtual sound.

The sound synthesis models are designed not only to be intuitively controllable by the user, but also to be computationally affordable for the machine. The desired efficiency is obtained through *cartoonification*, a specific design constraint implying a simplification of the physical descriptions and a consequent reduction of the available synthesis parameters. This economy of means exaggerates the most salient timbral aspects of the virtual sound events, a desired side effect which ultimately leads to a higher perceptual clarity of the simulation.

As previously mentioned in Section 2.1, the SDT sound models are used as basic building blocks to compose *timbral families*, categories of imitated sounds that are un-

ambiguously discriminable in terms of interaction, temporal and timbral properties. Whether composed by one or more low-level synthesis models, a timbral family is described in terms of specific, appropriate spaces and trajectories of sound synthesis parameters. All the timbral families (i.e., the blue boxes in Figure 1) are implemented and made available as Max patches in the current release of the toolkit.

2.3 A tool for sketching sonic interactions

Being temporary and disposable communication devices, sketches need to be produced with little time and effort. The more the resources required to produce a sketch, the greater the risk of being unwilling to throw it away in favor of possibly better options. The main advantage offered by drawn sketches in the early stages of a visual design process is the possibility to quickly materialize, store, compare and iteratively refine different ideas, gradually moving from early intuitions towards working prototypes.

The cartoonified, computationally affordable models of the SDT attempt to afford the same kind of interaction in the acoustic domain, enabling the sketching of sonic interactions in real-time on ordinary hardware, with a tight coupling between sound synthesis and physical objects to be sonified. The comparison and refinement of sonic sketches are made possible by means of saving and recalling presets of synthesis parameters. Presets can be further edited on GUIs or with MIDI/OSC external devices.

The almost direct relationship between synthesis parameters and basic physics facilitates understanding and creativity in sound design, supporting the unfolding of the designer's intentions on synthetic acoustic phenomena that are readily available and accessible through the concept of timbral family. Efforts are focused on providing economical control layers and parameter spaces, to interpret and control the physical descriptions of sound events in an intuitive way.

3. SKAT-STUDIO

SkAT-Studio is a prototype demonstration framework designed to facilitate the integration of other Max technologies in vocal and gestural sonic sketching.

3.1 Application workflow

A SkAT-Studio configuration is composed of the five following stages:

Input: Acquisition of voice and gesture;

Analysis: Extraction of meaningful features and descriptors from the input;

Mapping: Transformation of the analysis features into synthesis parameters by further elaboration, rescaling and/or combination;

Synthesis: Production of sound. This can be either purely procedural sound synthesis or post-processing of an existing sound (e.g. pitch shifting or time stretching);

Output: Playback or recording of the final sound.

3.2 Software overview

The framework is designed with flexibility and modularity in mind, and it is entirely developed in Max. It is composed by a main GUI (see Figure 2) which can host and link together a collection of loadable *modules*, each one taking care of a specific operation in the global process. Several modules can be loaded simultaneously, and signal and/or control data can be routed at will among different modules using *patchbays*.

Many different modules can be loaded at any given time, leading to possible cluttering of the interface and computing performance issues. To mitigate this problems, each of the five stages (input, analysis, mapping, synthesis, output) is materialized as a *group*. Groups help organizing information and simplifying the use of the software. Each group may contain several modules, whose control data and signals can be routed to other modules in the same group or even to an external group. Each SkAT-Studio module belongs to a group, according to its function.

A wide variety of modules is already available in the framework, offering the basic building blocks for the composition of complex configurations. The acquisition of audio signals from a microphone (input), the extraction of one feature or a set of features (analysis), the linear transformation of a parameter (mapping), the implementation of sound models (synthesis) and the direct playback through the speakers (output) are just some of the functions offered by the SkAT-Studio core modules.

In addition, users can easily build and add their own modules inside the SkAT-Studio framework. Each module is realized as a separate Max patch, which must adhere to a simple module template. The template provides a common interface for back-end communication with the other parts of the framework and front-end integration into the main GUI. To comply with the template, modules must graphically fit a given area, and provide the following information:

- Name of the module,
- Inputs and outputs of the module (number and names),
- Documentation (input/output data types, author, description of the underlying algorithms and so on).

The interactive and visual nature of the Max patching environment, combined with the simple yet versatile module template, allows quick and easy integration of new features into the system.

Audio signals and control data can be freely routed from any output of a module to any input of any other module, using routing matrices called *patchbays*. A patchbay is a double entry table, as displayed in Figure 3, with all the module outputs listed on the top row and all the module inputs listed on the left column. A toggle matrix allows to associate each output to one or more inputs, simply activating the appropriate toggles in the double entry table.

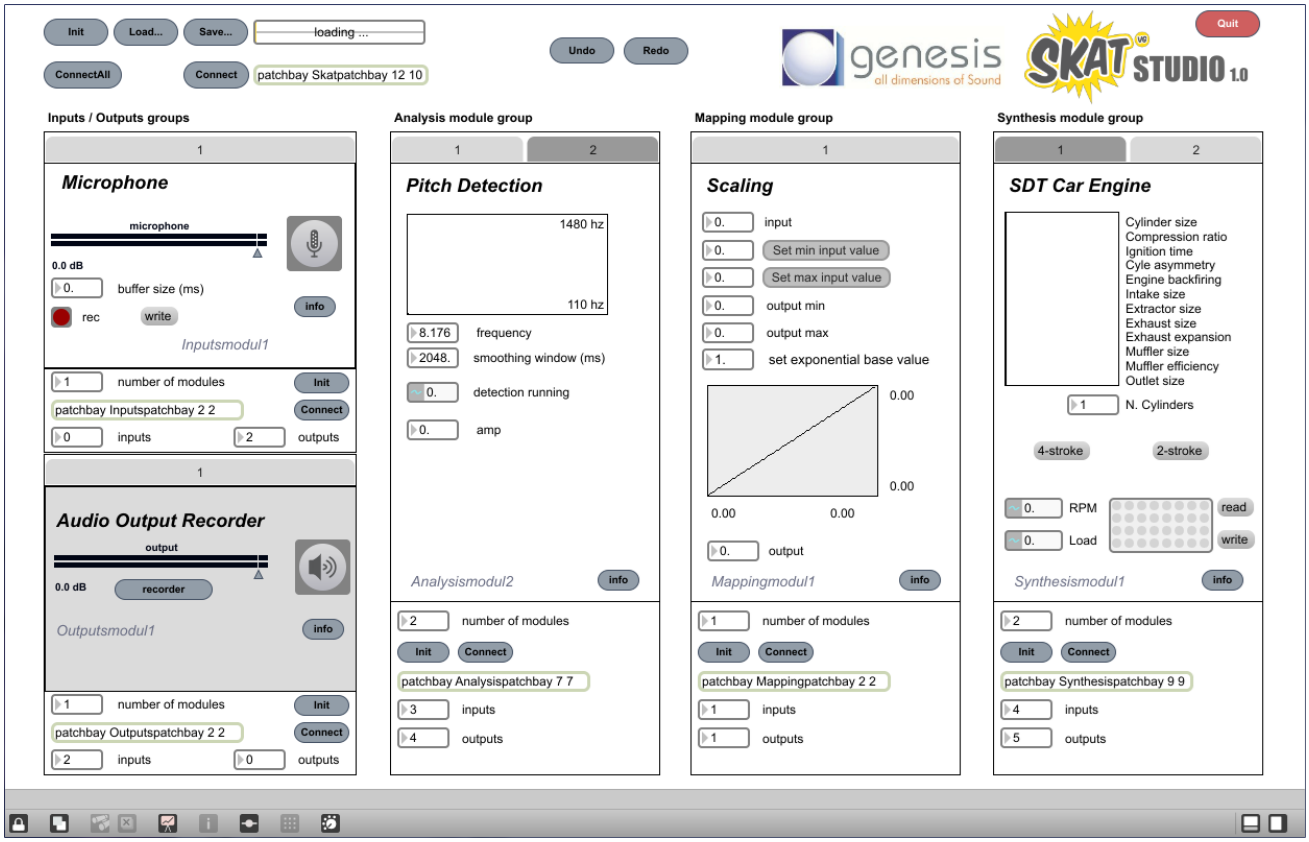

Figure 2. The SkAT-Studio workflow.

The subdivision of modules into functional groups, originally introduced to reduce conceptual and interface clutter, also simplifies the data routing process. Inbound and outbound data are first routed among modules inside a group, and successively among the groups inside the main framework. The framework therefore includes six patchbays: One for each group, plus a global one for the whole system.

3.3 Building a configuration

SkAT-Studio configurations can be built by performing a series of simple operations through the application GUI. The first step is choosing how many modules need to be loaded in each group. This operation creates as many tabs as required in the corresponding canvases. Modules can then be loaded in the tabs, either by drag and drop from a file manager or by choosing the module from the list visualized in the empty tab.

The next step is defining the number of inputs and outputs that each group should expose to the global routing patchbay of the system. By default, it is the total number of inputs/outputs of all the modules instantiated in the group. However, avoiding to expose data which do not need to go outside of the group allows to reduce the amount of routing connections, and therefore the size of the global patchbay.

Once everything is set up in place, the last step consists in clicking on the *connect* buttons to open the patchbays and route data inside of each group and among different groups. Once the configuration is ready, the sound designer can work with it and produce sonic sketches by tweaking the parameters exposed by the different modules. The possibility to save and load timbral family presets, together with an undo/redo history function, allows to compare, refine and possibly merge different sketches.

4. IMITATION-DRIVEN SOUND SYNTHESIS

The expressive power of human voice and gestures can be exploited to control the sound synthesis process and leveraged to perform quick and rough explorations of the parameters space of the available algorithms, shaping sound by mimicking the desired result. Taken together, the Sound Design Toolkit and SkAT-Studio provide an integrated environment for imitation-driven sound synthesis, in which sound designers can go from concepts, through exploration and mocking-up, to prototyping in sonic interaction design, taking advantage of all the possibilities offered by vocal and gestural imitations in every step of the process. The global workflow of the system is composed of two steps:

Select: The user produces a vocal imitation of the desired sound. The vocal imitation is recognized, classified, and the corresponding timbral family and vocal/gestural control layer are selected.

Play: The user controls the synthesizers in real time using vocalization and gesture, navigating the timbral space of the selected model and iteratively refining her sonic sketches. The use of voice and gesture to

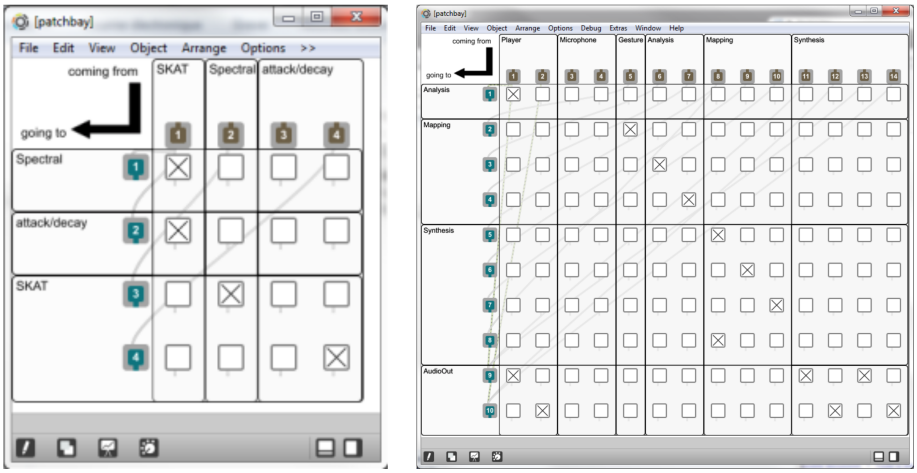

Figure 3. On the left, example of a patchbay for the analysis group. On the right, the global SkAT-Studio patchbay.

control sound production allows a fast, direct and easy manipulation of the synthesis parameters.

The *Select* step accepts a sound signal as input, and outputs a SkAT-Studio configuration which defines the behavior of the *Play* step.

The first step towards imitation-driven sound synthesis is the extraction of meaningful information from the vocal signal, in the form of higher level features and descriptors. To accomplish this task, the SDT has been enriched with tools for the analysis of audio signals in addition to the collection of sound synthesizers. A wide range of well-documented audio descriptors [25–27], have been reimplemented and made available as SDT externals. Recent studies on vocal imitations of basic auditory features and identification of sound events, however, pointed out that effective imitation strategies for the communication of sonic concepts exploit a few and simple acoustic features, and that the features cannot be consistently and reliably controlled all together, at the same time [28, 29].

In this respect, only a limited amount of descriptors is actually useful, and an even smaller subset is used to control a timbral family at any given time. Voice and gesture are used for a coarse control of the synthesis models, leaving further timbral refinement to manual operation on the graphical user interface or other external devices [12]. When placed on the visual canvas of the Max patcher, and connected in a coherent data flow, the SDT components can be operated via GUI sliders and knobs or external devices to refine the result. The extraction of features for control purposes includes:

- Amplitude variations and temporal patterns;
- Fundamental frequency, closely related to the sensation of pitch;
- Signal zero crossing rate, a rough estimate of the noisiness of a sound;
- Spectral centroid, directly related to the sensation of brightness of a sound;
- Spectral energy distribution, changing for different vowels.

Each of the SDT analysis externals is embedded in a separate SkAT-Studio module, to allow its inclusion in SkAT-Studio configurations.

The descriptors obtained by the analysis modules must then be mapped to the synthesis parameters of the available models and used to control the temporal behavior of the sound models. For each timbral family, a small subset of the available descriptors is scaled, combined and assigned to the vocally controlled synthesis parameters. All the operations involved in this process are performed by SkAT-Studio modules belonging to the mapping group.

At this stage, a simple, yet effective set of control maps per timbral family has been devised, which meets the listener expectations about the behavior of the sound producing events. For example, as the energy of an impact is expected to affect the amplitude and the spectral bandwidth of the resulting sound, similarly the timbral characteristics of its imitation will produce the same effect. In other words, it is possible to exploit the common relations between timbral features and physical parameters. Some examples include:

- The *pitch* of a vocal signal can be directly mapped to the revolutions per minute of both combustion engines and electric motors;
- The spectral *centroid* can be related to the concept of size (for instance, the size of bubbles in liquid sounds);
- The spectral *spread* can be associated to the concept of hollow body resonance, as found in many timbral families (e.g cavities in an air flow, the chassis of an electric motor, the exhaust system of a combustion engine, a container filled with a liquid, etc.);
- The temporal and spectral *onset* information can be used to trigger discrete events, like single impacts or explosions;

- The *zero crossing rate* of a vocal imitation can be put in relation with the graininess in higher level textures such as rolling, rubbing, scraping and crumpling, to the harshness of machine sounds, and in general to all the synthesis parameters related to the concept of noisiness.

Finally, the output of the mapping modules is routed to the synthesis group, to generate the sonic sketch. Each timbral family defined in the SDT is ported into SkAT-Studio as a synthesis module, exposing the vocally controlled synthesis parameters as inputs and the generated audio signal as output. Although not all the timbral possibilities provided by the synthesis modules are reproducible and controllable by vocal imitations, it is nevertheless possible to produce convincing and recognizable sonic sketches by mimicking a few salient, perceptually-relevant features for their identification. More subtle nuances, not directly controllable by vocal input, can be tweaked on the GUI of each module using traditional input methods such as virtual sliders and knobs.

To summarize, the proposed framework strives to facilitate the sound designer by providing models of sounds that humans can think of and represent through their voice and gestures. This aspect is reflected in the general procedural audio approach informing the SDT algorithms, and in the organization of SkAT-Studio workflows and configurations.

5. CONCLUSIONS AND FUTURE WORK

Although not fully evaluated yet, SDT and SkAT-Studio have been successfully used together for sketching the sonic behavior of a driving simulator, in the context of virtual reality and augmented environments [30]. Imitation-driven sound synthesis has also been presented and used in a series of sound design workshops, conducted as part of the SkAT-VG project.

We recently involved expert sound designers, in the 48 Hours Sound Design workshop [3] at Chateau La Coste art park and vineyard, in south France. 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. Vocal sketching methods and tools (including SkAT Studio and SDT) were the exclusive means available for sound ideas generation and sketching. In general, the technological support to vocal production and sketching was positively received, as the sound designers managed to explore and produce a large set of sounds in a very limited set of time. Yet, the provided SDT palette of sound models was found to be too bounded to realistic behaviors. The sound designers were also concerned about the cartoonified quality of the resulting sound. However, this rather reflected their inclination to produce well-refined sound propositions from the very beginning of their creative process, thus stressing a certain reluctance towards sketching and its purpose.

[3] The documentary of the workshop is available at: `https://vimeo.com/169521601`.

Indeed, vocal sketching in cooperative sound design tasks have been extensively documented, during a recent workshop held in November 2015 at the Medialogy course of Aalborg University Copenhagen, Denmark, and it is currently undergoing a process of detailed protocol and linkographic analyses [31]. Protocol and linkographic analyses are aimed at producing a fine-grained understanding of the cognitive behaviors in sound design tasks, measure the efficiency of the creative process, and ultimately assess the effectiveness of vocal sketching methods. Hence, the design of the sketching tools is grounded in the development of skills and practices of sound representations.

Acknowledgments

The authors are pursuing this research as part of the project SkAT-VG and acknowledge 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.

6. REFERENCES

[1] K. Franinović and S. Serafin, *Sonic interaction design: fresh perspectives*. Mit Press, 2013.

[2] D. Rocchesso, "Sounding objects in europe," *The New Soundtrack*, vol. 4, no. 2, pp. 157–164, 2014.

[3] S. Delle Monache and D. Rocchesso, "Bauhaus legacy in research through design: The case of basic sonic interaction design," *International Journal of Design*, vol. 8, no. 3, pp. 139–154, 2014.

[4] D. Hug and N. Misdariis, "Towards a conceptual framework to integrate designerly and scientific sound design methods," in *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, ser. AM '11. New York, NY, USA: ACM, 2011, pp. 23–30. [Online]. Available: http://doi.acm.org/10.1145/2095667.2095671

[5] S. Greenberg, S. Carpendale, N. Marquardt, and B. Buxton, *Sketching user experiences: The workbook*. Boston: Morgan Kaufmann, 2012.

[6] G. Lemaitre and D. Rocchesso, "On the effectiveness of vocal imitations and verbal descriptions of sounds," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 862–873, 2014.

[7] 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*, vol. 138, no. 3, pp. 1780–1780, 2015.

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

[9] B. Caramiaux, A. Altavilla, S. G. Pobiner, and A. Tanaka, "Form follows sound: Designing interactions from sonic memories," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 3943–3952. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702515

[10] C. Erkut, S. Serafin, M. Hoby, and J. Sårde, "Product sound design: Form, function, and experience," in *Proceedings of the Audio Mostly 2015 on Interaction With Sound*, ser. AM '15. New York, NY, USA: ACM, 2015, pp. 10:1–10:6. [Online]. Available: http://doi.acm.org/10.1145/2814895.2814920

[11] D. Rocchesso, G. Lemaitre, P. Susini, S. Ternström, and P. Boussard, "Sketching sound with voice and gesture," *Interactions*, vol. 22, no. 1, pp. 38–41, 2015.

[12] D. Rocchesso, D. Mauro, and S. Delle Monache, "miMic: the microphone as a pencil," in *Proceedings of the Tenth International Conference on Tangible, Embedded and Embodied Interaction*, ser. TEI '16. ACM, 2016.

[13] J. Janer, "Singing-driven interfaces for sound synthesizers," Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, 2008.

[14] S. Fasciani and L. Wyse, "A voice interface for sound generators: adaptive and automatic mapping of gestures to sound." in *Proceedings of the 12th International Conference on New Interfaces for Musical Expression (NIME)*, 2012.

[15] M. M. Wanderley and P. Depalle, "Gestural control of sound synthesis," *Proceedings of the IEEE*, vol. 92, no. 4, pp. 632–644, 2004.

[16] W. W. Gaver, "The sonicfinder: An interface that uses auditory icons," *Hum.-Comput. Interact.*, vol. 4, no. 1, pp. 67–94, Mar. 1989. [Online]. Available: http://dx.doi.org/10.1207/s15327051hci0401_3

[17] M. Rath and D. Rocchesso, "Continuous sonic feedback from a rolling ball," *IEEE MultiMedia*, vol. 12, no. 2, pp. 60–69, 2005.

[18] W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[19] S. Delle Monache, P. Polotti, and D. Rocchesso, "A toolkit for explorations in sonic interaction design," in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, ser. AM '10. New York, NY, USA: ACM, 2010, pp. 1:1–1:7. [Online]. Available: http://doi.acm.org/10.1145/1859799.1859800

[20] 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*, vol. 16, no. 1, p. 16, 2010.

[21] O. Houix, G. Lemaitre, N. Misdariis, P. Susini, and I. Urdapilleta, "A lexical analysis of environmental sound categories." *Journal of Experimental Psychology: Applied*, vol. 18, no. 1, p. 52, 2012.

[22] G. Lemaitre, F. Voisin, H. Scurto, O. Houix, P. Susini, N. Misdariis, and F. Bevilacqua, "A large set of vocal and gestural imitations," SkAT-VG Project, Tech. Rep., November 2015, deliverable D4.4.1. [Online]. Available: http://skatvg.iuav.it/wp-content/uploads/2015/11/SkATVGDeliverableD4.4.1.pdf

[23] A. Farnell, "Behaviour, structure and causality in procedural audio," in *Game sound technology and player interaction concepts and developments*, M. Grimshaw, Ed. New York, NY, USA: Information Science Reference, 2010, pp. 313–329.

[24] D. Svanæs, "Understanding interactivity: Steps to a phenomenology of human-computer interaction, monograph," Ph.D. dissertation, NTNU, Trondheim, Norway, 2000. [Online]. Available: http://dag.idi.ntnu.no/interactivity.pdf

[25] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *Journal of the Acoustical Society of America*, vol. 130, no. 5, p. 2902, 2011.

[26] P. McLeod and G. Wyvill, "A smarter way to find pitch," in *Proceedings of International Computer Music Conference, ICMC*, 2005.

[27] D. Stowell and M. Plumbley, "Adaptive whitening for improved real-time audio onset detection," in *Proceedings of the International Computer Music Conference (ICMC 07)*, Copenhagen, Denmark, 2007.

[28] G. Lemaitre, A. Dessein, P. Susini, and K. Aura, "Vocal imitations and the identification of sound events." *Ecological psychology*, vol. 23, no. 4, pp. 267–307, 2011.

[29] G. Lemaitre, A. Jabbari, O. Houix, N. Misdariis, and P. Susini, "Vocal imitations of basic auditory features," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 2268–2268, 2015.

[30] S. Baldan, H. Lachambre, S. D. Monache, and P. Boussard, "Physically informed car engine sound synthesis for virtual and augmented environments," in *Proceedings of the 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*. IEEE, 2015.

[31] S. Delle Monache and D. Rocchesso, "Cooperative sound design: A protocol analysis," in *Accepted for publication in Proc. of AudioMostly 2016, a conference on interaction with sound*, Norrköping, Sweden, 2016.