FP7-ICT-2013-C FET-Future Emerging Technologies-618067

SkAT-VG:
Sketching Audio Technologies using Vocalizations and Gestures

D7.7.1
Interactive prototypes realized with the SkAT-VG tool

First Author	Stefano Delle Monache
Responsible Partner	Genesis
Status-Version:	Final-1.1
Date:	January 9, 2017
EC Distribution:	Consortium
Project Number:	618067
Project Title:	Sketching Audio Technologies using Vocalizations and Gestures

Title of Deliverable:	Interactive prototypes realized with the SkAT-VG tool
Date of delivery to the EC:	09/01/2017

Workpackage responsible for the Deliverable	WP7
Editor(s):	-
Contributor(s):	Stefano Baldan, Patrick Boussard, Alan Del Piccolo, Stefano Delle Monache, Clément Dendievel, Hélène Lachambre, Davide Rocchesso
Reviewer(s):	Davide Rocchesso
Approved by:	All Partners

Abstract	This deliverable presents the SkAT-VG prototypes of sonic interactive vocal sketching tools
Keyword List:	Sonic interaction design

The SkAT-VG Consortium consists of the following entities:

#	Participant Name	Short-Name	Role	Country
1	Università Iuav di Venezia	IUAV	Co-ordinator	Italy
2	Institut de Recherche et de Coordination Acoustique/Musique	IRCAM	Contractor	France
3	Kungliga Tekniska Högskolan	KTH	Contractor	Sweden
4	Genesis SA	GENESIS	Contractor	France

Document Revision History

Deliverable D7.7.1

Version	Date	Description	Author
v1	2016/12/12	Draft the outline, import documents. Polish everything	SDM
v2	2017/01/09	Completed	SDM

Table of Contents

Index of Figures

List of Acronyms and Abbreviations

DoW Description of Work

EC European Commission

PM Person Month

WP Work Package

GA Grant Agreement

CA Consortium Agreement

M Milestone

Mo Month

Q Quarter

1 Content of this deliverable

1.1 Deliverable 7.7.1 within WP7

Work package 7 "WP7 - Sonic Interaction Design" of the SkAT-VG project studies the use and exploitation of vocal utterances and gestures from the design viewpoint. In WP7 basic research and scientific findings are distilled in relevant and systematized vocal and gestural sketching practices. In this respect, the role of WP7 is to collect scientific knowledge and software development to inform the design research and assessment of the emerging sound design tools and methods. More importantly WP7 acts as a hub, a dispatcher, by providing design feedbacks that are reflected in the experiments and studies of the other WPs, thus establishing a shared methodological connection between scientific and design research.

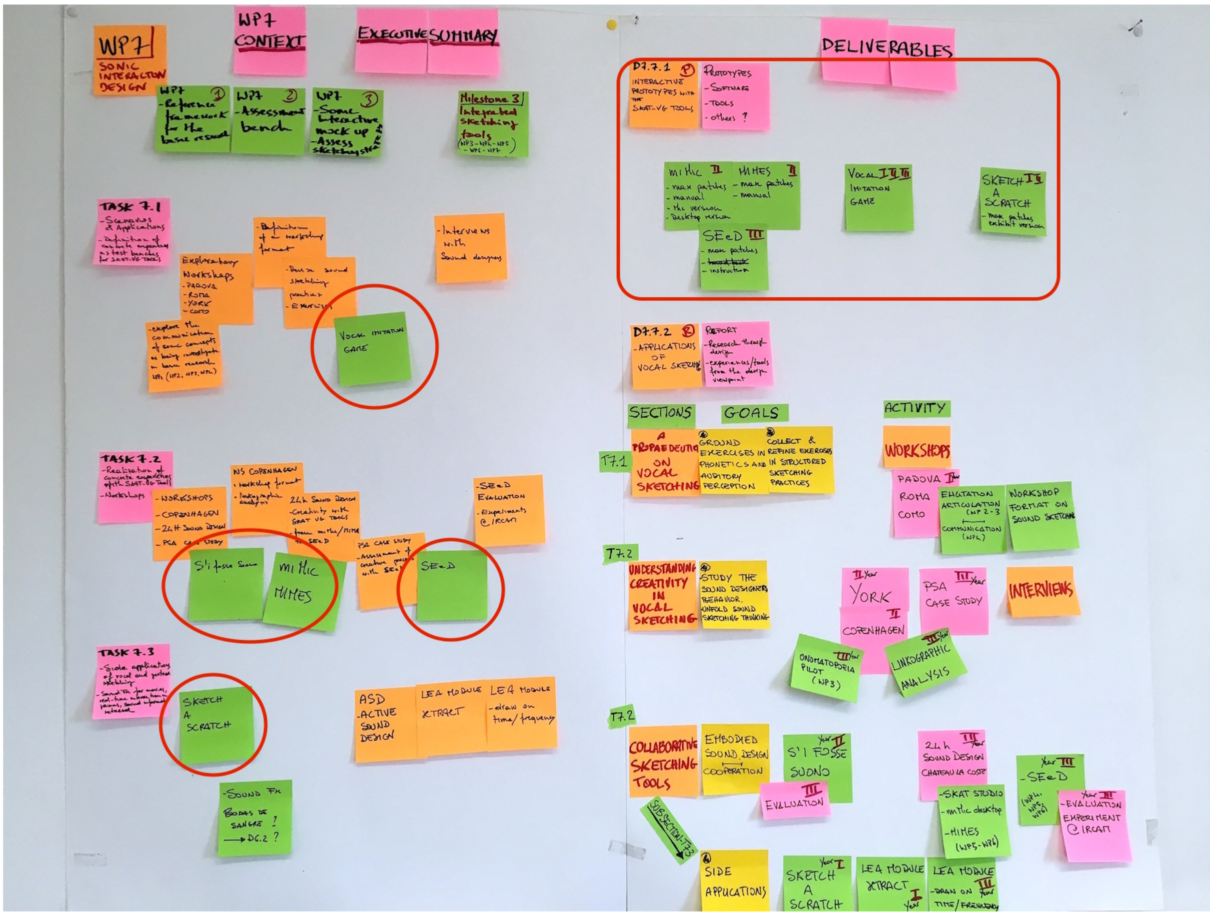

Figure 1: WP7 context, map of the research activities per task.

As depicted in the planning board in Figure 1, the main objectives of WP7 are:

1. To provide a reference framework based on sonic interaction design (SID) cases to nurture the basic research in the other WPs (task T7.1);

2. To provide an experimental assessment bench of the SkAT-VG results, based on the implementation of sonic interactive mock-ups and demonstrators (task T7.2);

3. To synthesize the project contributions in design tools and methods to test the overall sketching strategy (task T7.2, and T7.3).

Deliverable D7.7.1 presents applications, software and tools of vocal sketching that have been developed in WP7, and accompanies the extended report on the WP7 design research activities, included in the SkAT-VG Deliverable D7.7.2. These applications emerged throughout the execution of the three tasks of WP7, and they are outlined in the ovals in Figure 1.

One main tool developed throughout research-through-design workshops is the *vocal imitation game*. Originally conceived as a tool to investigate the communication of vocal imitations, it progressed towards a structured guessing game, that showed to be an effective means to foster the practice of vocal sketching.

SkAT Studio, *miMic*, and *MIMES* are the three SkAT-VG demonstrators that emerged throughout tight collaborations between WP7, WP6, and WP5. They have already been included in the SkAT-VG Deliverables D6.6.1 and D6.6.2. In this deliverable we essentially release the Cycling'74 Max patches of miMic and MIMES. The three demonstrators have been extensively used in workshops, and were eventually merged in the software architecture of SEeD, which represents the main contribution of this deliverable. For the in-depth discussion of SEeD we refer to Deliveable D7.7.2.

2 The Vocal Imitation Game

The Vocal Imitation Game is an educational tool in the form of a guessing game, suitable for small teams of players. The collaborative game is based on a card deck of verbal descriptions of sounding objects. Guessing games are largely used to reinforce concepts in (children) education and to foster experiential learning. In addition, decks of cards are often used as tools to support learning and decision-making in design education and practice[1]

The goal is to guess the sound-producing action or object mimicked with the voice. Each player/performer at turn chooses a card from the pack of verbal descriptions of objects/interactions, mimics the sound of that object, and tries to guide the other teammates towards the right guess.

The game develops analytical skills through listening, and guessing constitutes an implicit evaluation of the performer's strategy in vocal imitation. The competitive aspect contributes through engagement in the iterative virtuoso exploration of the vocal capabilities, while at the same time the collaborative environment fosters social interaction and collaboration among peers. To progress in the game, it is necessary to enact continuous negotiations with the other players through the development and practice of one own's vocal abilities

The vocal imitation game was originally conceived as a research tool to elicit the production of vocalizations to communicate sonic concepts, aimed at investigating the behavioral aspects and emerging strategies in vocal and gestural imitation. In parallel, the game has been constantly improved with respect to the card deck, rules and pacing. The thorough discussion of the design research activity carried out through the vocal imitation game, and its exploitation in the SkAT-VG workshops is reported in the SkAT-VG Deliverable D7.7.2.

2.1 The deck of cards

The current deck of cards is formed by a list of sound-producing actions and objects, which refer to machine sounds (e.g., button and switches, hums, air conditioning, vehicles, food processors, and wipers) and mechanical interactions involving liquids, solids, and gases (without focusing

[1] `http://www.methodkit.com/research-method-cards/`, a comprehensive overview of existing types of packs for design purposes.

on the sources), reported in Table 1. The ambiguity of referent sounds is exploited to increase engagement, and to foster virtuosity and competitiveness in the game.

The layout of the deck of cards, together with the game rules, can be downloaded at the following link: `http://skatvg.iuav.it/wp-content/uploads/2014/02/SkAT-VG-Vocal_Imitation_Game-1.zip`.

Machines		**Mechanical interactions**
appliances→aerodynamic:**air conditioner**	mechanisms→car equipment: **windshield wipers**	gas: **crackling noise**
appliances→aerodynamic: **hair dryer**	mechanisms→car equipment: **electric window up**	gas: **gunshot**
appliances→aerodynamic: **vacuum cleaner**	mechanisms→car equipment: **turning signal**	gas: **fireworks**
appliances→aerodynamic: **electric fan**	mechanisms→car equipment: **electric window down**	gas: **spraying**
appliances→no motor: **boiling water**	mechanisms→doors: **shutting the door**	liquids: **water burbling**
appliances→no motor: **moka coffee brewing**	mechanisms→doors: **creaky door**	liquids: **filling a glass with water**
appliances→no motor: **gas stove**	mechanisms→doors: **squeaking hinges**	liquids: **water sloshing**
appliances→no motor: **pepper grinder**	mechanisms→doors: **knocking on door**	liquids: **water dripping**
appliances→office: **typrewriter**	mechanisms→transportation: **car idling**	solids: **crumpling a can**
appliances→office: **computer printer**	mechanisms→transportation: **bus idling**	solids: **duct tape**
appliances→office: **typing on keyboard**	mechanisms→transportation: **tractor idling**	solids: **scribbling**
appliances→office: **photocopier**	mechanisms→transportation: **motorbike idling**	solids: **sharpening blades**
appliances→with motor: **blender**	mechanisms→alarm bells: **electronic door bell**	
appliances→with motor: **electric shaver**	mechanisms→alarm bells: **disarm car alarm**	
appliances→with motor: **fridge**	mechanisms→alarm bells: **alarm clock**	
appliances→with motor: **washing machine**	mechanisms→alarm bells: **bicycle bell**	
mechanisms→button-switches: **striker wheel (lighter)**	mechanisms→hand tools no motor: **filing**	
mechanisms→button-switches: **computer key**	mechanisms→hand tools no motor: **sawing**	
mechanisms→button-switches: **light switch**	mechanisms→hand tools no motor: **sanding**	
mechanisms→button-switches: **shutter button**	mechanisms→hand tools no motor: **hammering**	
	mechanisms→hand tools with motor: **drilling**	
	mechanisms→hand tools with motor: **electric screw driver**	

Table 1: The list of referent sounds populating the deck of cards in the vocal imitation game.

3 Collaborative and Embodied Sound Design Tools

Within the SkAT-VG project, WP6 and WP7 led to the development of three main demonstrators, which have been extensively tested and used in design research activities and workshops [BDRH16, HDL+16].

3.1 SkAT Studio

SkAT Studio, shown in Figure 2, is a modular framework based on the Cycling'74 Max environment and conceived by GENESIS to host the different technologies developed by the other partners of the project. SkAT Studio was one of the main tools used in the "48 hours of sound design" workshop that took place in Spring 2016, at Château La Coste, a vineyard and art center located in Le Puy-Sainte-Réparade, in southern France. SkAT Studio

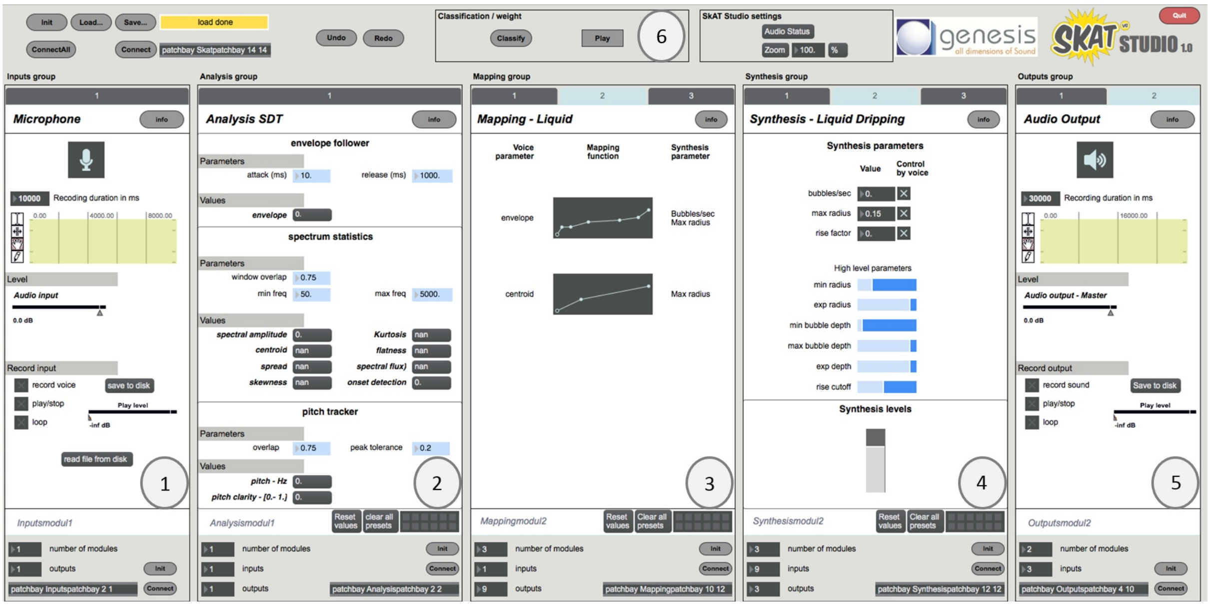

Figure 2: SkAT Studio in action. The control scenario for the blowing sound model.

is extensively described and released as prototype in Deliverable D6.6.1. The software can be downloaded as Max package from the SkAT-VG Github repository at the following link: `https://github.com/SkAT-VG/SkATStudio/archive/V1.0.zip`.

3.2 MIMES

MIMES is another Cycling'74 Max prototype developed by IRCAM. In the MIMES system, shown in Figure 3, vocal sound ideas are automatically analyzed and acquired as temporal and morphological profiles of acoustic features (loudness, spectral centroid, spectral skewness, spectral spread, noisiness and zero-crossing rate). The resulting list of features is used as input parameters to drive corpus-based sound synthesis [Sch07], a technique similar to real-time audio mosaicing [SSCL10].

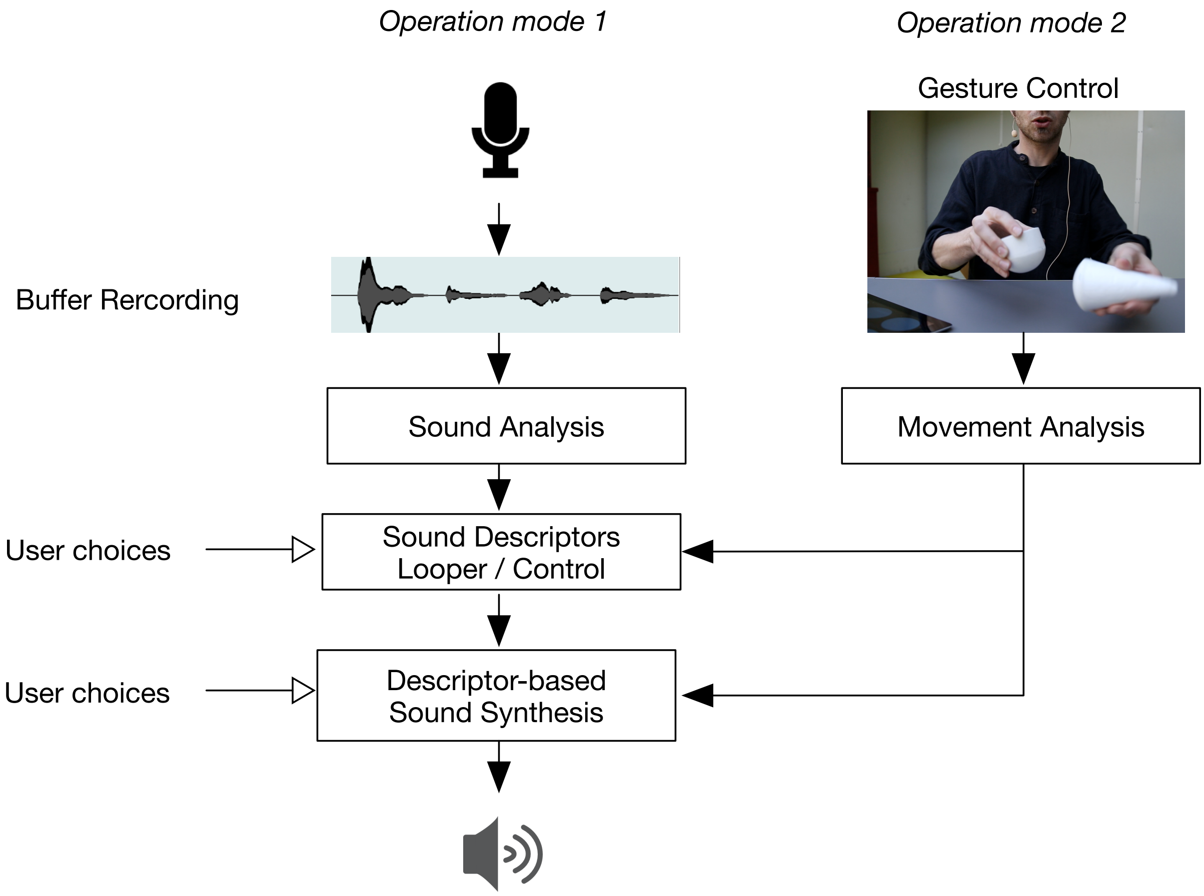

Figure 3: MIMES software architecture.

The software description and documentation is available in Deliverable D6.6.2. The software has been recently improved and released within WP7. It can be downloaded at the following link: `http://skatvg.iuav.it/wp-content/uploads/2014/02/SkAT-VG-MIMES-V3.zip`.

3.3 miMic

miMic is the third SkAT-VG demonstrator, developed by IUAV. It was first shown at the ICT2015[2], Lisbon, on 20-22 October, 2015. miMic is an eyes-free interface for vocal and gestural sketching, based on a microphone augmented with embedded inertial sensors and two buttons for contextual functions (i.e., train, set, play modes). The Cycling'74 Max program handles the classification of the vocal imitations as instances of a palette of physics-based sound models, which can be set up and controlled in real-time [RMM16]. miMic, shown in Figure 4, acts like a pencil in the hand of the sound designer, who can vocalize a sound idea and convert it into a mixture of synthetic sound models. An extended description of the software architecture has been reported in the SkAT-VG Deliverable D6.6.2. A desktop version of miMic, shown in Figure 5 was used during the "48 hours of sound design"

[2] `http://ec.europa.eu/digital-agenda/events/cf/ict2015/item-display.cfm?id=14872`.

Figure 4: miMic, the microphone as a pencil

workshop. The miMic demonstrator constituted the basis for the fusion with MIMES towards

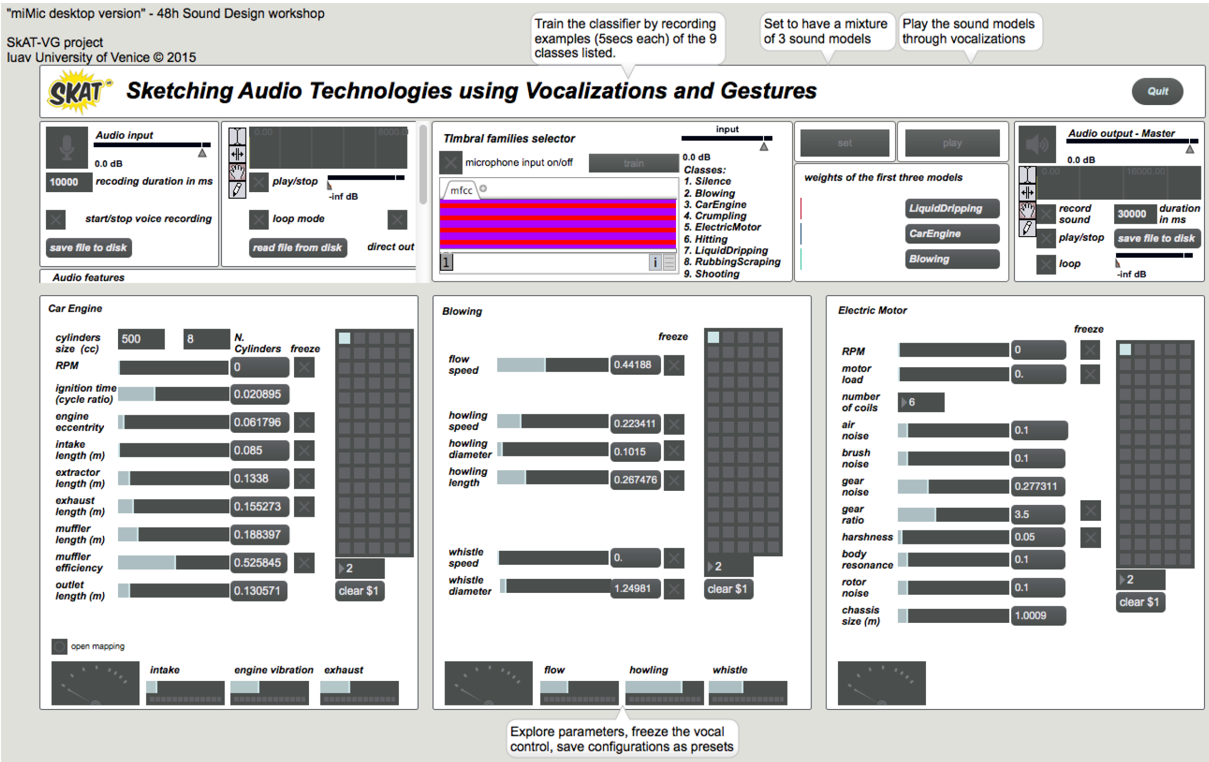

Figure 5: miMic desktop with GUI.

the development of SEeD, based on the recommendations collected in the debrief meeting with the sound designers (see SkAT-VG Deliverable D7.7.2). The Max program of the two versions of miMic can be downloaded at:
`http://skatvg.iuav.it/wp-content/uploads/2014/02/SkAT-VG-miMic_V1.zip` - microphone version

http://skatvg.iuav.it/wp-content/uploads/2014/02/SkAT-VG-miMicDesktop_V2.zip - desktop version.

3.4 SEeD - Sonic Embodied Design

The software architecture of SEeD emerged as a fusion of the two approaches to sonic sketching embodied in miMic and MIMES. SEeD represents the most advanced interactive prototype towards the achievement of an automatic system that translates vocal and gestural signals into synthetic sonic sketches.

The diagram of Sonic Embodied Design is depicted in Figure 6. Two main modes are available, *Set* and *Play*, a feature borrowed from the previous architecture of miMic [RMM16] and extended to include both physics-based and corpus-based sound synthesis models (see SkAT-VG Deliverables D6.6.1 and D6.6.2). The experience collected through the development of the proof-of-concept artistic installation *S'i fosse suono*[3] with the sound designer Andrea Cera convinced us to keep both approaches, as they respectively match the reported concrete and acousmatic attitudes of the artist–designer to synthetic sound.

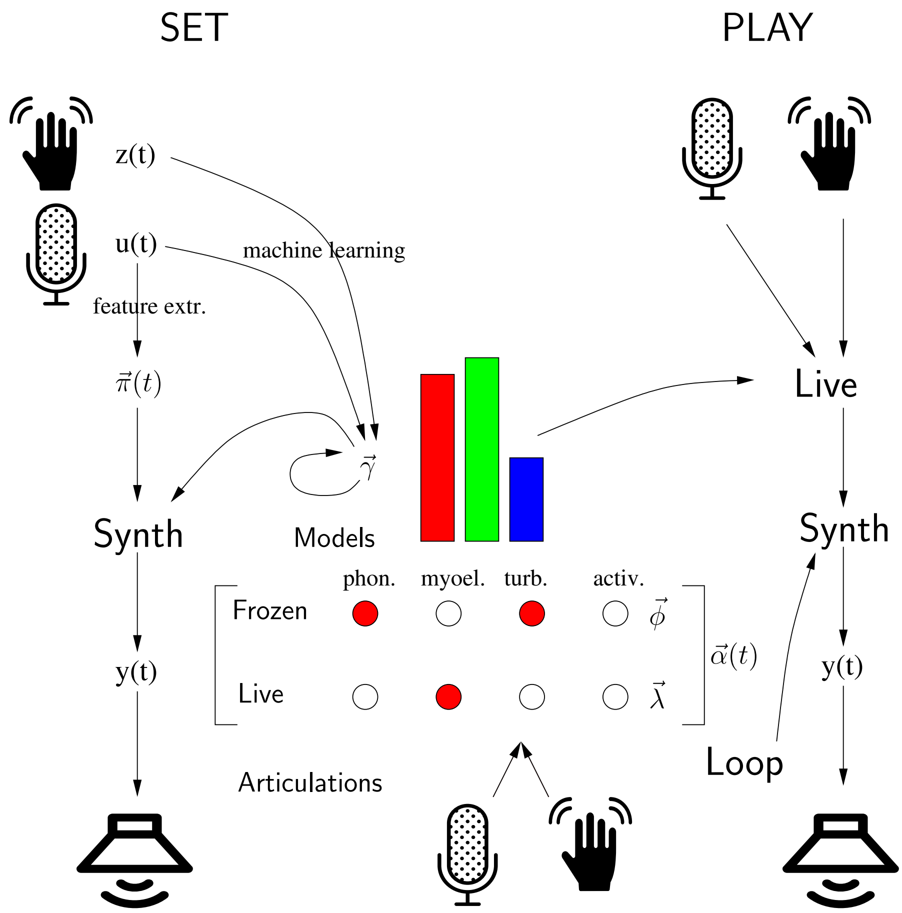

Figure 6: Sketch of SEeD

[3]In its web version, and without multi-touch support, *S'i' fosse suono* is available at skatvg.eu/SIFosse/

Vocal and gestural signals are captured through a microphone and an inertial measurement unit. In the *Set* mode, imitations of referent sounds are automatically classified into sound categories by a machine learning algorithm, and analyzed by a series of timbral features extractors. The system returns a vector of weights (colored bars) and a vector of functions of time, which drive the selection and temporal control of a subset of sound synthesis models to generate an output signal similar to the input utterance. The relative weights of the subset of models can be further adjusted manually by the user.

Once the sound models are selected, the *Play* mode empowers the user with a direct vocal and gestural control of their temporal behavior. In the case of physical models, each model is provided with a control layer which performs a real-time mapping of the vocal and gestural features extracted from the input signals into synthesis parameters. In the case of corpus-based synthesis, the same vocal and gestural features are used to select and concatenate short fragments of the sound corpora, trying to match as closely as possible the timbral and morphological properties of the input [MP15]. Alternatively, one or more sound synthesis parameters can be looped by exploiting the vectors of weights and features returned in the *Set* mode.

1. Download and installation

The software can be downloaded at `http://skatvg.iuav.it/wp-content/uploads/2014/02/SkAT-VG-SEeD.zip`. SEeD runs on Cycling'74 Max 7 and Mac OS X only.

- To install the program, uncompress the zip file and locate it within the file preferences path of Max;
- The SDT-v076 [BDR17] and MubuForMax-1.9.1 [SRS+09] are located as zip files in the *external* folder. Uncompress the files and copy them in `Documents/Max 7/Packages`;
- Double-click on `SkAT-VG-SEeD-v1.maxpat` to launch SEeD.

Figure 7 shows the main elements of the user interface which mirror the diagram sketch. The patch is organized into three main sections, as highlighted by the blue-framed boxes. The leftmost box corresponds to the topmost part of SEeD diagram:

(1) `concat` (for corpus-based) and `physmod` (for physics-based) synthesis techniques can be dynamically switched throughout set and play modes;

(2) *Set* mode and *Play* mode are shared by both concat and physmod. This means that once a mixture of sound classes is set, their relative weights $\vec{\gamma}$ affect either the physmod sound models or the concat sound corpora. In play mode, the *Live* submode allows to control the sound synthesis through vocalizations and gestures, while in the *Loop* submode the sound models are driven by accessing the stream $\vec{\pi}$ of audio and motion descriptors previously recorded in the set mode (and displayed in the `voicedescriptors` tab). Additionally, in the *Loop* submode the stream $\vec{\pi}$ can be replaced by a new recording, to support the exploration of the sonic space of the sound models/corpora set, through creative utterances. The new stream does not affect the previously-set selection and weighting.

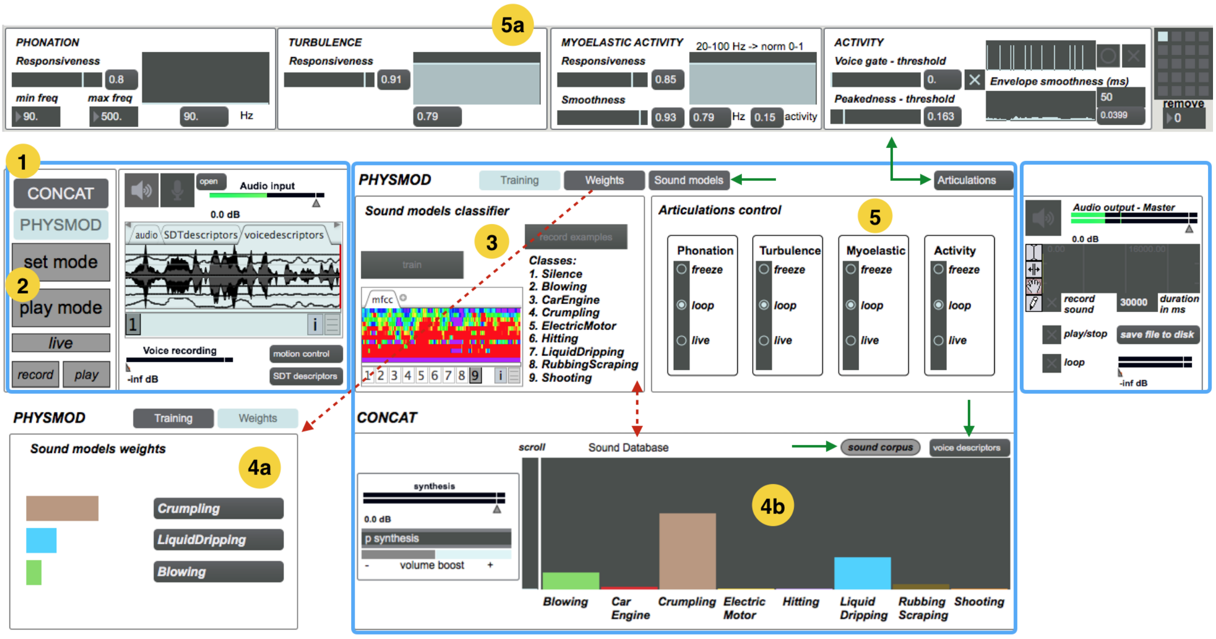

Figure 7: SEeD: elements of graphical interface

The central blue-framed box shows the high-level user interface of the sound synthesis engine. For the sake of clarity, both `physmod` and `concat` interfaces are shown together in Figure 7, although in the patch they are displayed alternatively according to the selected synthesis technique:

(3) The user trains the classifier used in the set mode (in this case a Gaussian Mixture Model) by providing vocal imitations of the available sound models/corpora. Eight classes of sound models are implemented, corresponding to those used for the realization of the sounds in *S'i' fosse suono*: blowing, car engine, crumpling, electric motor, hitting, liquid dripping, rubbing scraping, and shooting;

(4a) The weights tab shows the best three sound models and their relative contributions. Eventually their balance in the mixture can be adjusted manually, by operating the sliders;

(4b) The mixture selected in the *Set* mode shapes the balance of the eight classes of sound samples used for the concatenative synthesis;

(5) The articulation control window implements the bottom central part of the SEeD diagram, in which the loop submode allows to `freeze`, `loop`, or act `live` on articulatory features;

(5a) The articulatory features are integrated to give a high-level description of vocalizations in terms of *phonation*, *turbulence*, *myoelasticity*, and general activity. This layer can be accessed to tailor the system responsiveness to one's own vocal characteristics.

The rightmost blue-framed box represents the output module of the chain, in which it is possible to record the sound session. Finally, the green arrows point towards low level interfaces to fine-tune the sound models and corpora configurations, together with their control maps.

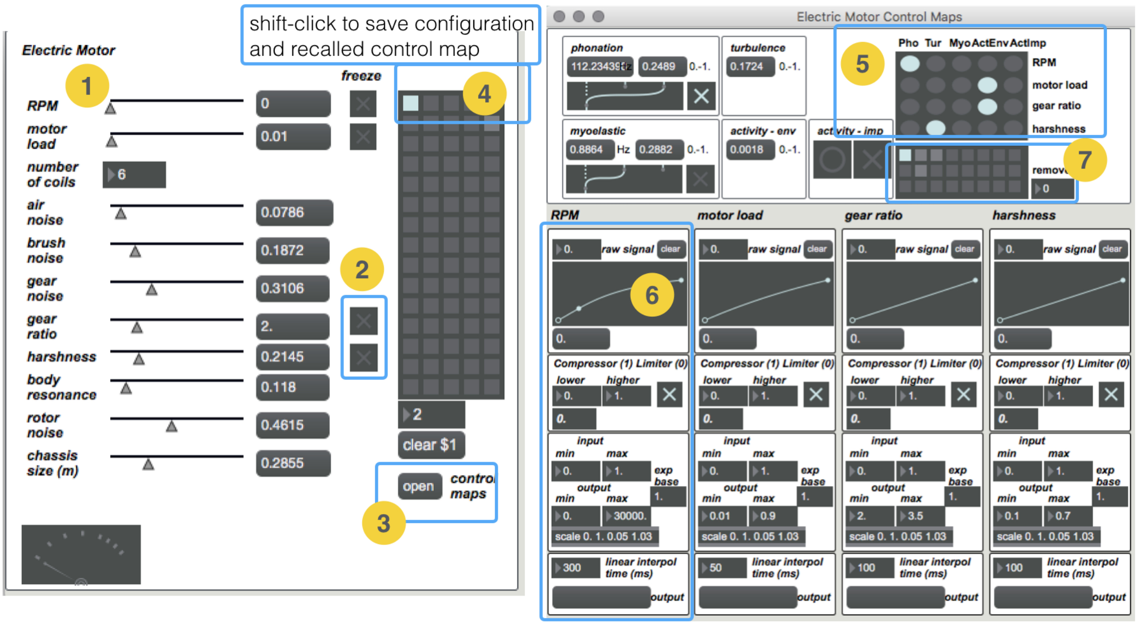

Figure 8: SEeD: elements of graphical interface, control maps.

Figure 8 shows an example of the sound models user interface and the corresponding control maps. The user interface can be accessed via the sound models tab in Figure 7 (4a):

(1,2) the GUI shows all the available control parameters of the sound model at hand (electric motor in the example). The control parameters of each sound model of the Sound Design Toolkit are extensively covered in Deliverable D6.6.1. The voice-driven parameters are provided with a *freeze* function which stops or resumes their continuous update;

(3) opens the corresponding control map;

(4) custom configurations of non interactive parameters can be saved as presets. The presets also store the currently loaded control map (see 7);

(5) phonation, turbulence, myoelasticity and activity can be associated to specific control parameters in a one-to-one routing;

(6) the mapping module per parameter allows to manipulate the incoming raw signal of the vocal articulation at hand. Signals can be compressed or limited, scaled and linearly interpolated;

(7) control map configurations can be saved as presets. These can be recalled directly from the preset in the main GUI of the sound model.

4 Side Applications: Sketch-a-Scratch

Sketch-a-Scratch is an abstract experimental workbench, shown in Figure 9, conceived to explore several research-through-design topics: Sonic sketching of surface qualities; creative texture modeling and multimodal exploration; exploration of auditory contents rendered by means of auditory, visual and tactile feedback [RDP16, DDR+15].

Figure 9: Installation configuration of Sketch-a-Scratch: Top of the installation with tablet, stylus and audio recorder. The rounded token with the red led allows to browse and switch between the four textures.

The Cycling'74 Max program, shown in Figure 10, can be downloaded at `http://skatvg.iuav.it/wp-content/uploads/2014/02/SkAT-VG-Sketch_a_Scratch.zip`.

Within the context of SkAT-VG Sketch-a-Scratch was used to investigate the production of virtual surface profiles based on vocal input, that could be explored through pen-based interaction, and further rendered through digital sound models and textures. The system is based on a graphic tablet with digital display, visual surface textures can be "scratched" with a stylus augmented with auditory and vibrotactile feedback. Sound and vibrations are rendered in real-time by the physically-informed impact and friction models of the Sound Design Toolkit (described in detail in the SkAT-VG Deliverable D6.6.1). The application was also instrumental to the early development of the Sound Design Toolkit in the project.

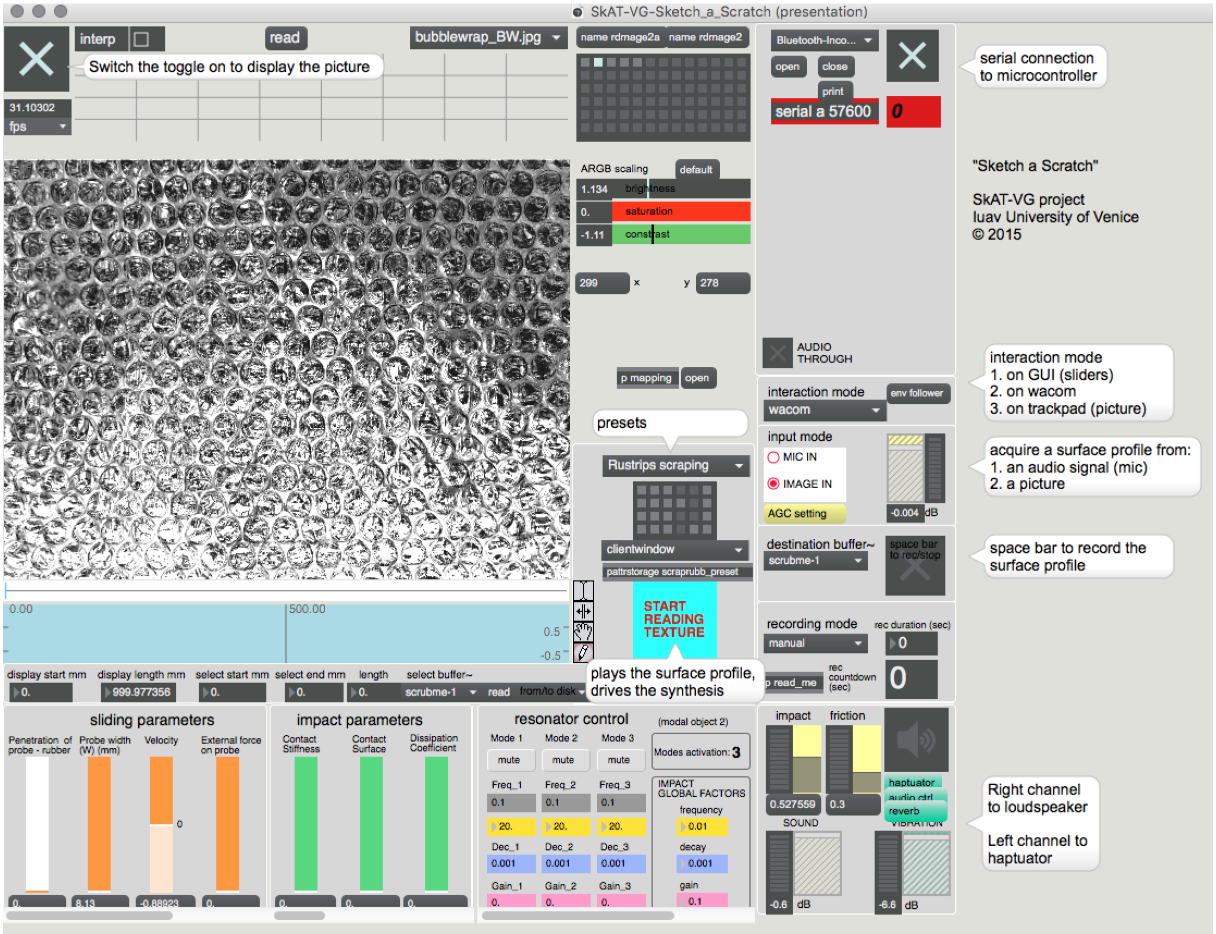

Figure 10: Sketch A Scratch GUI.

References

[BDR17] Stefano Baldan, Stefano Delle Monache, and Davide Rocchesso. The sound design toolkit. *SoftwareX*, 2017. In press.

[BDRH16] Stefano Baldan, Stefano Delle Monache, Davide Rocchesso, and Lachambre Helene. Sketching sonic interactions by imitation-driven sound synthesis. In *Proc. 13th Sound and Music Computing Conference*, SMC 2016, pages 47–54, 2016.

[DDR+15] A Del Piccolo, S Delle Monache, D Rocchesso, S Papetti, and Davide Andrea Mauro. To "sketch-a-scratch". In *Proc. 12th Sound and Music Computing Conference*, Maynooth, Ireland, 2015.

[HDL+16] Olivier Houix, Stefano Delle Monache, Hélène Lachambre, Frédéric Bevilacqua, Davide Rocchesso, and Guillaume Lemaitre. Innovative tools for sound sketching combining vocalizations and gestures. In *Proceedings of the Audio Mostly 2016*, AM '16, pages 12–19, New York, NY, USA, 2016. ACM.

[MP15] Enrico Marchetto and Geoffroy Peeters. A set of audio features for the morphological description of vocal imitations. In *Proc. of the 18th Intl. Conf. on Digital Audio Effects*, 2015.

[RDP16] D Rocchesso, S Delle Monache, and S Papetti. Multisensory texture exploration at the tip of the pen. *International Journal of Human-Computer Studies*, 85:47–56, 2016.

[RMM16] Davide Rocchesso, Davide A Mauro, and Stefano Delle Monache. mimic: The microphone as a pencil. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 357–364. ACM, 2016.

[Sch07] Diemo Schwarz. Corpus-based concatenative synthesis. *IEEE Signal Processing Magazine*, 24(2):92–104, 2007.

[SRS+09] Norbert Schnell, Axel Röbel, Diemo Schwarz, Geoffroy Peeters, and Riccardo 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)*, Montreal, Canada, August 2009.

[SSCL10] Norbert Schnell, Marco Antonio Suarez-Cifuentes, and Jean-Philippe Lambert. First Steps in Relaxed Real-Time Typo-Morphological Audio Analysis/Synthesis. In *Proceedings of the Sound and Music Computing conference*, Barcelone, Spain, 2010.