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

SkAT-VG: Sketching Audio Technologies using Vocalizations and Gestures

D1.1.3 Final Report

First Author	Davide Rocchesso
Responsible Partner	IUAV
Status-Version:	Draft-1.0
Date:	February 10, 2017
EC Distribution:	Consortium
Project Number:	618067
Project Title:	Sketching Audio Technologies using Vocalizations and Gestures

Title of Deliverable:	Final Report
Date of delivery to the EC:	28/02/2017

Workpackage responsible for the Deliverable	WP1
Editor(s):	Davide Rocchesso
Contributor(s):	Davide Rocchesso, Stefano Delle Monache, Sten Ternström, Guillaume Lemaitre, Olivier Houix, Patrick Susini, Nicolas Misdariis, Geoffroy Peeters, Patrick Boussard, Hélène Lachambre, Clément Dendievel
Reviewer(s):	Davide Rocchesso
Approved by:	All Partners

Abstract	This Final Report comprises: a A final publishable summary report; b A plan for the use and dissemination of foreground; c A report covering the wider societal implications of the project.
Keyword List:	final report; final publishable summary report; plan for the use and dissemination of foreground; societal implications

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 D1.1.3

Version	**Date**	**Description**	**Author**
Final	10/02/2017		ROC
First draft	03/11/2016	Import from template	ROC
First draft	01/12/2016	A few updates (WP4)	GL

Project Final Report

Grant Agreement number: FP7-618067

Project acronym: SkAT-VG

Project title: Sketching Audio Technologies using Vocalizations and Gestures

Funding Scheme: FP7-ICT-2013-C FET-Future Emerging Technologies

Date of latest version of Annex I against which the assessment will be made: 30/08/2013

Period covered: from 01 January 2014 to 31 December 2016

Name, title and organisation of the scientific representative of the project's coordinator:

Prof. Davide Rocchesso,
Dipartimento di Culture del Progetto,
Università Iuav di Venezia
Tel: +39 041 257 1852
E-mail: roc@iuav.it

Project website address: http://www.skatvg.eu

Contents

1 Final publishable summary report

1.1 Executive summary

Pleasant, yet functional. The fusion of these two adjectives describes most of the work of designers, in any domain. In the sound domain, designers aim at giving a pleasant and functional 'voice' to the objects that will populate future soundscapes. Improved safety, health, and quality of life are the possible benefits for society at large. The SkAT-VG project has been looking for ways to make the sound design process easier and more effective. The fundamental idea was to exploit the most natural tools for sound ideation: Human voice and gestures.

Humans have surprising capabilities in communicating sound, but a thorough understanding of how this happens and could be exploited required an ambitious research plan: The SkAT-VG project. Three research institutions and one company joined their skills in perception, voice production, gestures, machine learning, sound synthesis, and interaction design to expand the scientific understanding in voice- and gesture-based sound communication, as well as to introduce new tools and methods for sketching in sound design processes.

One major outcome of SkAT-VG has been to scrutinize the possibilities and limitations of human voice and gestures as sound sketching apparatuses, by collecting multi-signal recordings of thousands of utterances (provided by both laypersons and voice professionals), and analyzing their relation to the sonic phenomena they mimic. Experimental studies have shown that vocal imitations generally have higher communicative power than verbal portraits of reference sounds. A vocal imitation is not a cheap version of the referent sound, but rather an optimized representation that emphasizes only those features that are important for identification. The analysis of recordings has also shown that accompanying gestures have mainly an iconic function, with gestural metaphors largely shared across population. In addition, the large collection of audiovisual recordings of vocal and gestural imitations (made available to the research community) offers the opportunity to further enquire how people perceive, represent, and communicate about sounds.

The second major outcome of the project has been a set of algorithms connecting vocal and gestural inputs to sound synthesis. The automatic extraction of the most relevant primitives of vocal and gestural imitations has been the *leitmotif* of SkAT-VG research in machine learning, where several classifiers have been developed by exploitation of low-level acoustic features, emerging time-frequency components, and phonetic annotation. Once a sound category has been automatically extracted from an imitation, a sound model is activated and controlled by voice and gesture to synthesize that class of sounds. SkAT-VG has used two complementary sound synthesis approaches: The simulation of a physical process deduced from the imitation (physical modeling), and the exploitation of extracted primitives to weight and control prerecorded audio corpora (concatenative synthesis).

The third major outcome of the project has been the development of sound design methodologies, guidelines, and tools. The SkAT-VG project has demonstrated that vocal/gestural sketching can be as immediate and effective for sound design as drawing on paper is for visual design. The project has developed sound-design prototype instruments, which allow the designer to transform such sketches into versatile sound models that can be adapted and manipulated. These developments have the potential to re-define current sound design practices by including embodied interaction and cooperation in the creative flow. Extensive interaction with professional and academic stakeholders has given valuable insights into how the research results can lead to product innovation. In the near future, designers might sketch novel responsive sounds, for a car or for a coffee maker, by using the whole expressive potential of their voice and body.

1.2 Project context and objectives

Our world is constantly making noise. Often the world is cluttered by too many noises, hardly distinguishable from each other and contributing to acoustic pollution. However, sound is vital to humans and most animals, as it brings information about events and processes that occur in nature, or about actions of humans and animals.

In traditional societies, sounds and noises were mainly byproducts of human actions, with different activities resulting in different "keynote" sounds. With the industrial revolution, the mechanization in production processes and in daily life increased the level of background noise in urban areas and made many subtle sounds largely irrelevant for human perception and action. In the post-industrial society, sound has been gaining back its role as a mediator of human actions and, at the same time, it has been contributing to the aesthetic character of everyday products, services, and activities. Most sounds that surround us are produced by human artefacts, and an increasingly larger part of them are explicitly introduced to augment, or accompany, or support human activities. As such, sounds are entities that must be designed, to make products and services contribute to a better quality of life.

Sound design certainly plays an important role in contemporary creativity culture, and it has many different application areas, including entertainment, gaming, and products such as cars or appliances. In many of these areas, sound is functional to effective and engaging interactions, and it needs to be designed as a dynamic entity that manifests itself in the context of an interaction. Sonic interaction design is the specialized design discipline and set of skills that should drive the work of creative workers in these areas.

Despite the urgent need of sound creators in many different domains, in most cases their work can be hardly classified as a design discipline. Sound design often lacks the most basic characteristics of design disciplines: It is solipsistic rather than cooperative; It refuses iteration through multiple cycles of evaluation and re-design; It lacks sketching practices. There are multiple reasons for this state of affairs, among these the lack of proper education programs, which for product and visual design have nurtured progress in knowledge and practice since the early twentieth century.

Perhaps the most striking difference between the work of sound designers and that of visual or product designers, is that the former seem to skip the sketching phase of the design process. Visual designers and architects are used to sketch with paper and pencil, or with digital sketching tools, in order to let their ideas emerge and develop. Sketching is the quick production of many design embryos, which can be rapidly compared and shared with peers and stakeholders. Sketches are food for thought, they are not good-looking cakes to be exhibited in a pastry shop. Conversely, in the domain of sound, designers most often go direct to complex software tools and large databases of samples and processing effects to find a refined sound to be shared with clients. This feed-forward process undermines creativity, as it easily leads to the production of cliché sounds, rather unnoticeable in the jungle of artificial sounds that pervades our environments. Even more serious is the difficulty of designing sounds that are suitable for interactivity, that respond dynamically and support action. For interactive applications, digital audio workstations and soundbanks are of little use.

Assuming that the introduction of sketching practices in sound design would be beneficial, where is the analogue of paper and pencil to be found? Indeed, humans do sound sketching since the early months of their life, and even as adults we sketch a sound when we cannot find the words to describe an event or process. In most of the cases, our voice is the "pencil" we exploit to convey sonic ideas, often accompanied by gestures. More than a "paper" to record these ideas, which could be trivially obtained by a voice recorder, sound designers would need a tool that interprets their vocalizations and produces articulations of dynamic sound models.

Such tool and synthesis models would allow to produce sound sketches rapidly, iteratively, and with the possibility of continuous dynamic adjustment. This is the missing enabling tool that would let sound designers do sketching in the early stage of the design process.

This is the vision of the SkAT-VG project, simple yet challenging. In fact, the effectiveness of vocal imitations as a means to communicate non-vocal sounds has been recognized and demonstrated only recently. A scientific understanding of why and how vocal and gestural imitations work is far from being complete, but investigations in this field are worth doing. In fact, vocal imitations are emerging as a sort of probe to study auditory perception at large. The voice gives us an embodied representation of sounds as perception-action ensembles, a physical anchor for the most salient sound elements. Furthermore, vocal imitations enable the communication of auditory experiences between two persons. As such, vocal imitations offer a window into listeners' cognitive representations of sounds, reflecting the content of sensory and memory representations of sounds.

Also, the production mechanisms and configurations involved in non-speech vocal communication have been largely left aside from scientific investigation, which has been previously polarized on speech and singing. Indeed, there are vocal emissions that are not found in any spoken language, but which are used in imitations, thus raising the need for new kinds of phonetic annotation. Overall, in order to serve the needs of a vocal sketching tool, the fundamental mechanisms of voice production in sound imitation need to be clearly identified. This can not be done without reference to linguistic articulations, as the boundary between sonic mimicry and linguistic aliases (onomatopoeia) is fuzzy, and humans seem to use a mixture of analog and symbolic utterances when communicating sounds. This is of course richly creative, and as listeners we enjoy being dazzled by skilful and elaborate imitations. But the emphasis in SkAT-VG has to be on first describing and then understanding 'mainstream' imitations of basic sounds, made without much prior training. For instance, do such imitations replicate the sounds as such, or do they caricature only the perceptually salient traits; or do the imitations allude to the actions causing the real sounds?

The connection between how vocal and gestural imitations are perceived as pointers to everyday sounds, and the articulatory mechanisms that are involved in such imitations, can be exploited via machine learning to automatically classify vocalizations and gestures to extract an appropriate sound model that represents what the given utterance is trying to convey. Given a set of examples belonging to different families of sound events, humans are good at mimicking such examples and at recognizing the events from their peers' imitations. By supervising a proper training by vocal imitations, a machine can be made to approximate human recognition performance, so to return a model that is appropriate for the imitated sound. A more general problem is that of producing and recognizing sound imitations from memory, without direct access to reference sound events. Humans can do that easily, but instructing a machine to do the same implies the identification of primitives that characterize the different classes of sound phenomena. The identification of primitives for both perception and production of vocal imitations helps selecting the features and orienting the classifiers, in a circle of correspondences between what is perceived, what is articulated, and raw signal features, as they are captured by microphones and sensors.

The practice of vocal sketching needs to be gradually introduced among sound designers. Education plays a major role, and learning activities can be effectively grounded in voice production mechanisms in sounds imitation. The emerging propaedeutics will structure methods to improve dexterity in vocal articulation and compositional skills, through continuous rehearsal. Design sessions and workshops, with both students and professionals, allow to better understand the design processes and to see how new practices and tools would affect them.

As soon as virtuosity sprouts, the access to one's own internal representations of sounds becomes easier, and a fine-grained sensory-motor control of vocal articulations opens the doors to inventiveness and control over the sound forms. Whenever thinking and reasoning in sound come into play, vocal sketching would endow practitioners with a powerful device to improve the sound design communication throughout their workflow, either with peers and with the other stakeholders longitudinally. In particular, sharing, embodiment and immediacy manifest through cooperation, and the study of cooperative sound design is a new endeavor that is revealing surprising behaviors. Formal methods to inspect the creative process of sound design are available to analyze the effectiveness of design tools such as the ones proposed by SkAT-VG. Embodied sound design tools are meant to foster immediate reflection, either individually or collaboratively, and fit the integrated workflow of software and tools, from the idea to the prototype. The creation and consolidation of a community of sound designers who collaborate with peers and stakeholders by using voice and gesture is as much a matter of technology as it is a matter of education and social practices. A research project such as SkAT-VG can be the seed that triggers such a cultural process.

In summary, the principal objectives of the SkAT-VG project are:

- To define and foster new research paths in perception and production of vocal imitations and expressive gestures;

- To develop automatic classifiers and extractors of primitives of vocal and gestural imitations, by integrating signal analysis with the physio-mechanics of vocal production;

- To explore the effectiveness of vocal and gestural sketching in sonic interaction design, by exploiting automatic classification for selection and parameterization of sound synthesis models.

1.3 Scientific and Technical Results

This section highlights the major outcomes of the SkAT-VG project, and refers to the project Work Packages (WP), as described in the Description of Work of the Grant Agreement. More details can be found in the reports for the respective deliverables, which are numbered by work package (e.g. D3.3.1 is the first deliverable of WP3) and available in the SkAT-VG website `www.skatvg.eu`.

1.3.1 A corpus of vocal and gestural imitations (WP2, WP3, WP4)

The study of vocal and gestural imitations of sounds being a relatively new field of investigation, no reliable resources were available at the beginning of the SkAT-VG project to base an analysis on. In particular, the repertoire of vocal and gestural articulations that are used to communicate non-speech and non-musical sounds was largely ignored. The construction of a corpus of reliable, controlled recordings has been a major endeavor of SkAT-VG, conducted in parallel by partners IRCAM and KTH. In Stockholm at KTH, a small number of voice professionals (actors) were recorded (see figure 1). In Paris at IRCAM, a large number of laypersons were recorded. At KTH the recordings included synchronized audio, video, and electroglottographic signals and were focused on the vocal apparatus. IRCAM, on the other hand, collected half-person audio-visual recordings, including gesture sensors (accelerometers) and 3D cameras. Participants imitated a large set of sounds, ranging from abstract sound effects to easily recognizable everyday sounds, using only their voice or combining vocal and gestural imitations. The two corpora are respectively called the KTH and the IRCAM datasets (D2.2.2).

Figure 1: Imitator participant as seen from the control room at KTH.

By analysis of the datasets, partner KTH developed an annotation system suitable for vocal imitations. The articulatory annotation is constituted by eight articulatory parameters that together give a holistic description of the articulations involved. The resulting annotation can be viewed as an articulatory score that describes the contribution or action of individual articulators over time. This annotation is rich in articulatory detail to ensure that all aspects that may be significant for conveying referent sounds through imitation are covered. The whole KTH dataset, and a selection of the IRCAM dataset were annotated, and the annotated recordings are available as tracks in the ELAN format[1] (D3.3.1). For making the unique data more accessible to the research community, the two corpora have been trimmed and co-organized and are being published on a common web site.

Multimodal Database of vocal and gestural imitations elicited by sounds The complete database is made accessbile from the web page
https://www.ircam.fr/project/blog/multimodal-database-of-vocal-and-gestural-imitations-elicited-by-sounds/.
It consists of three complementary databases:

1. A database (IRCAM) of vocal and gestural imitations of 52 referent sounds produced by French lay imitators. The referent sounds are mechanical sounds (12), basic physical interactions (12) and abstract sounds (10). The database is also composed by video recordings, sensor data and acoustical descriptors of each imitation. The database is structured as repertories of families of referent sounds and their associated imitations, classified by imitators;

[1] https://tla.mpi.nl/tools/tla-tools/elan/

2. A database (IRCAM) of vocal imitations made by a subset of French lay imitators and described in terms of the phonatory and articulatory configurations annotated by experienced phoneticians (KTH). The database is structured as continuous file of the 52 imitations for each imitator and the associated ELAN file;

3. A multimodal database (KTH) of vocal imitations of 50 referent sounds made by Swedish-speaking professional actors, described in terms of the phonatory and articulatory configurations employed for each of the sound productions. These configurations were assessed and annotated by experienced phoneticians, from recordings of the high-fidelity audio, dual-camera video, and electroglottographic (EGG) recordings. The database is structured as continuous file of the 50 imitations for each professional actor and the associate ELAN file.

For each database, an explanatory text details the structure and the different files.

Terms of use of the database Persons who want to download the recordings must first:

- register with an institutional e-mail address;
- accept these clauses concerning recordings:
  - use the data only within the framework of their research or within an educational framework;
  - do not use for any artistic or commercial purpose;
  - make no copies other than for the research work;
  - always cite the source of the database in each publication or presentation referring to this database. Reference: "Database of vocal and gestural imitations elicited by sounds, Copyright KTH & Ircam, created with the financial support of the Future and Emerging Technologies (FET) program within the Seventh Framework Program for Research of the European Commission under FET-Open grant number: 618067".

1.3.2 Production and phonetic representation of vocal imitations (WP3)

The articulatory annotation adopted in SkAT-VG makes use of eight annotation layers, or articulatory primitives, each representing a specific articulatory parameter. Two of the layers describe articulatory actions in the larynx, three layers describe the actions of the tongue (both tongue body and tongue tip), one layer is devoted to the lips, one layer controls for nasality and, finally, one layer describes the airstream mechanism (or sound initiation). A reannotation test was performed to assess the reliability of the annotation: Most of the articulatory parameters are robust to reannotation, but some are more subjective (D3.3.2). Figure 2 shows an example recorded fragment with annotation.

An inescapable aspect of voice-based communication of sound is that of onomatopoeia, which are linguistic aliases and, as such, culture dependent. In some cases it is difficult to tell if a vocalization is a true sound imitation or it is rather para-linguistic. In order to assess the relevance of onomatopoeia and its language dependence, an experiment based on a guess game was devised at KTH. The player must recognize linguistic-like dummy sounds and environmental sounds based on the indications given by an informant. The goal was to see whether participants will behave differently when they are in a linguistic or non linguistic

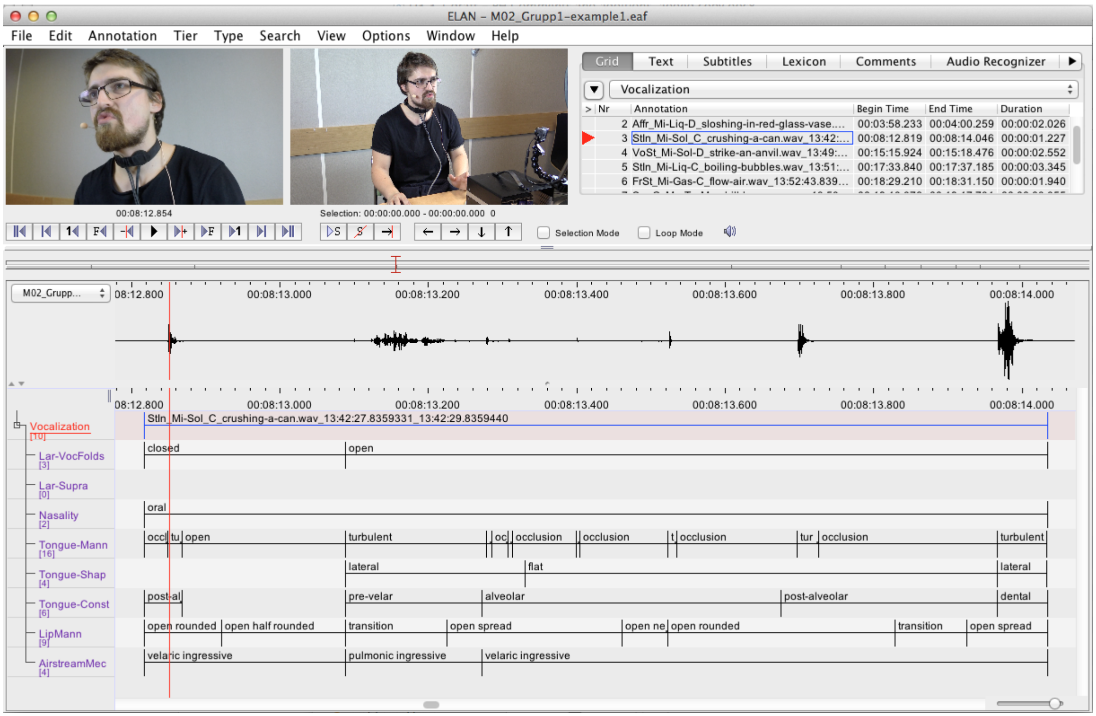

Figure 2: ELAN screenshot with a complex annotation example. The participant was imitating a crushing can.

context. The experiment showed that imitators are more inclined to use onomatopoeia when the recipient is a human who speaks their own language, and that the morphology of referent sounds strongly affects the imitation preferences (D3.3.3). A fairly clear distinction emerged between a 'linguistic' and a 'sonic' mode of imitation, the latter being preferred when the listener was understood to be a computer. So, linguistic aliases would probably not be of much help in an automated sound design system. However, this also raised the possibility of actively eliciting 'sound' mode imitation, for example, when training SkAT-VG users, or even to help students of pronunciation in a foreign language to bypass a native accent.

1.3.3 Perception and cognition of vocalizations and expressive gestures (WP4)

A strong foundation for the SkAT-VG project came from the preliminary experimental observation that vocal imitations communicate sounds more effectively than verbal portraits (Lemaitre and Rocchesso, JASA 135:2, 2014).

To study the perception of vocal imitations, the consortium had to identify a set of reference sounds that are unambiguously identified by human listeners. This led to the definition of three families and 26 categories. This selection included ten sounds of machines (e.g. a printer), ten sounds of mechanical interactions (e.g. the sound of tapping a board), and six abstract sounds (e.g. alarms), thus providing a good coverage of products sounds typically addressed by sound designers.

In order to understand how vocal imitations work in interpersonal sound communication, IRCAM studied how two expert and two lay participants reproduced four basic auditory features (pitch, tempo, sharpness and onset) of modulated narrow-band noises and pure tones, based on one feature or crossing two of the four features. It emerged that vocal imitations do not simply mimic the referent sounds, but seek to emphasize the characteristic features, within the constraints of human vocal production (D4.4.2; Lemaitre et al., JASA 139:1, 2016).

Vocal imitations may in fact be based on three strategies: faithfully reproducing the absolute value of the feature (e.g. pitch); Transposing the feature into the participants' registers (i.e. sharpness); Categorizing a continuum into discrete symbols (e.g. onsets).

A central question of the SkAT-VG project is how we can make sense of a vocal imitation when we hear it alone, i.e. relying on internal representations of different sonic categories, and not on the similarity between an imitation and the referent sound. For this purpose, IRCAM studied how well listeners can identify vocal imitations of basic mechanical interactions and machine sounds, without hearing the referent sounds. Human recognition performance of imitations was compared to recognition of sparsified time-frequency representations (auditory sketches) of the referent sounds, and it turned out that it is possible to produce vocal imitations that are identified almost as well as the referent sounds themselves, and not worse than moderately sparse representations (for most sounds). This study revealed that successful imitations are made of some caricatural rendering of the referent sounds, that may be actually different from the referent sounds, but still convey the idea of the features they refer to (D4.4.2; Lemaitre et al., PLoS ONE 11:12, 2016). This confirmed that vocal imitations are not "degraded" versions of the sounds they imitate, but optimized representations of the sounds, emphasizing only the features important for identification.

Another important issue is what is the role of gestures during imitations of sounds. Experiments were run with participants imitating different types of referent sounds with their voice and their hands. The results show that vocalizations reproduce more precisely than gestures the temporal aspects of the referent sounds (D4.4.2). Vocalizations can also reproduce more complex rhythmical patterns, whereas hand gestures seem to mainly tap a pulsation. The function of gestures is rather iconic, and the study identified several gestural metaphors shared across participants. For instance, most imitators represent pitch with a vertical metaphor. More surprisingly, participants shake their hands whenever they imitate a stable noisy texture, and use smoother gestures for tonal sounds or to indicate a temporal evolution of the spectrum. Even when imitators produce a shaky gesture to signify signals fluctuation, the regularity of the shaking matches the regularity of the signal. At least some individuals are also capable of using both voice and gestures to convey different layers of information, conveying one layer with their hand and another layer with their voice. These results provide the design work packages with important information, and they show that vocalizations and gestures have to be considered differently.

1.3.4 Analysis and classification of vocal and gestural imitations (WP5)

Within the SKAT-VG project, IRCAM has developed a system that allows the automatic recognition of the sound categories described in section 1.3.3 (alarms, wipers, fridge, printers, dripping, gushing, etc.) based solely on their vocal imitations. For this purpose, various strategies have been adopted (D5.5.1): from continuous/discrete Hidden Markov Models (HMM), to the development of dedicated morphological descriptors (that allow describing the specific temporal shape of each sound category), to the use of Dynamic Time Warping (DTW) for the alignment of sequences of audio features (see Figure 3). This led to the development of a system for automatic classification of imitations which would allow to launch the most appropriate synthesizer in an imitation-based sketching system.

By taking an articulatory viewpoint, KTH developed a set of analysis tools aiming at representing a vocal imitation as a set of vocal primitives (namely the turbulent, myoelastic, and phonetic components). These tools use the recently developed Auditory Receptive Fields (D5.5.1, Lindeberg and Friberg, PLoS ONE 10:3, 2015) analysis in conjunction with Partial

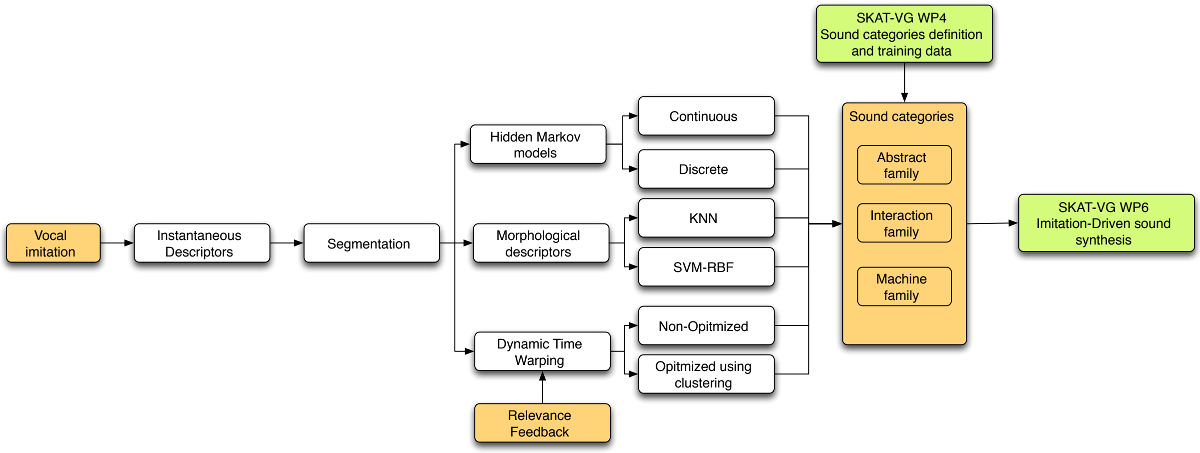

Figure 3: Systems to recognize sound categories from their vocal imitations.

Least Square Regression, and can be exploited for automatic transcription of vocal recordings into time series of vocal primitives, for informed classification of imitations (see Figure 4).

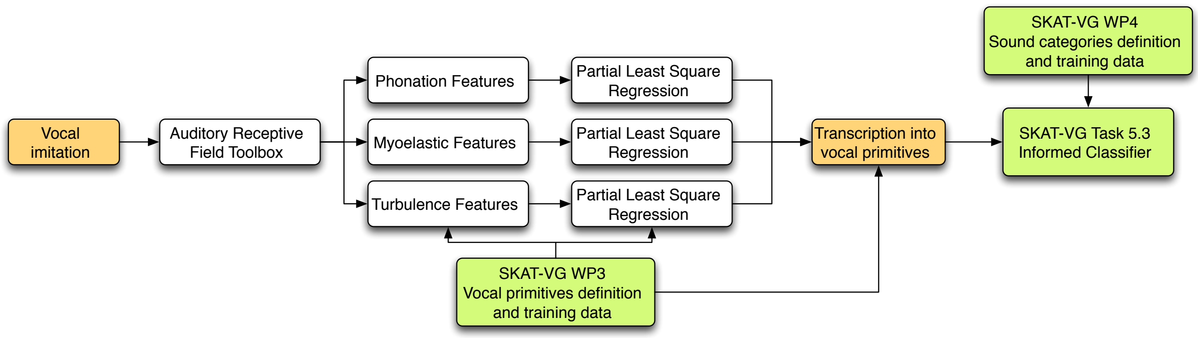

Figure 4: System to represent vocal imitations by automatic transcription into vocal primitives.

IRCAM also developed an innovative and promising system to automatically define and extract the set of audio primitives (time-frequency kernels) that allow to generate a given dataset of recordings of vocal imitations (D5.5.2). The method is based on an extension of Shift-Invariant Probabilistic Latent Component Analysis applied to dataset. These primitives can lead to the automatic annotation of tracks. They are then used to construct a set of HMMs specific to each sound category (see Figure 5).

The sound categories can also be partially described by the gestural signals that accompany vocal imitations. It appeared from our studies that the gesture dynamics, i.e. the signal frequency content, contains the most important and consistant information in this context. For this purpose, a novel movement representation of gestures, based on wavelet analysis has been developed at IRCAM (D5.5.2). It allows for extracting gestural primitives (steady, smooth, dynamic, impulse, periodic, shaky) that were found significant for gestural imitation of sound. First, the continuous wavelet transform is applied to signals provided by inertial measurement units (e.g. accelerometer). Second, the analysis of the scalogram allows for deriving motion descriptors that can be used by a classifier. Two complementary methods have been developed for the computation of these descriptors. The first one is based on particle filtering, (see Figure 6) which can track in real-time the evolution of a specific feature of the scalogram. The second one is based on Non-negative Matrix Factorization (NMF) of the scalogram (over a fixed window). We found that the NMF components, learned in a

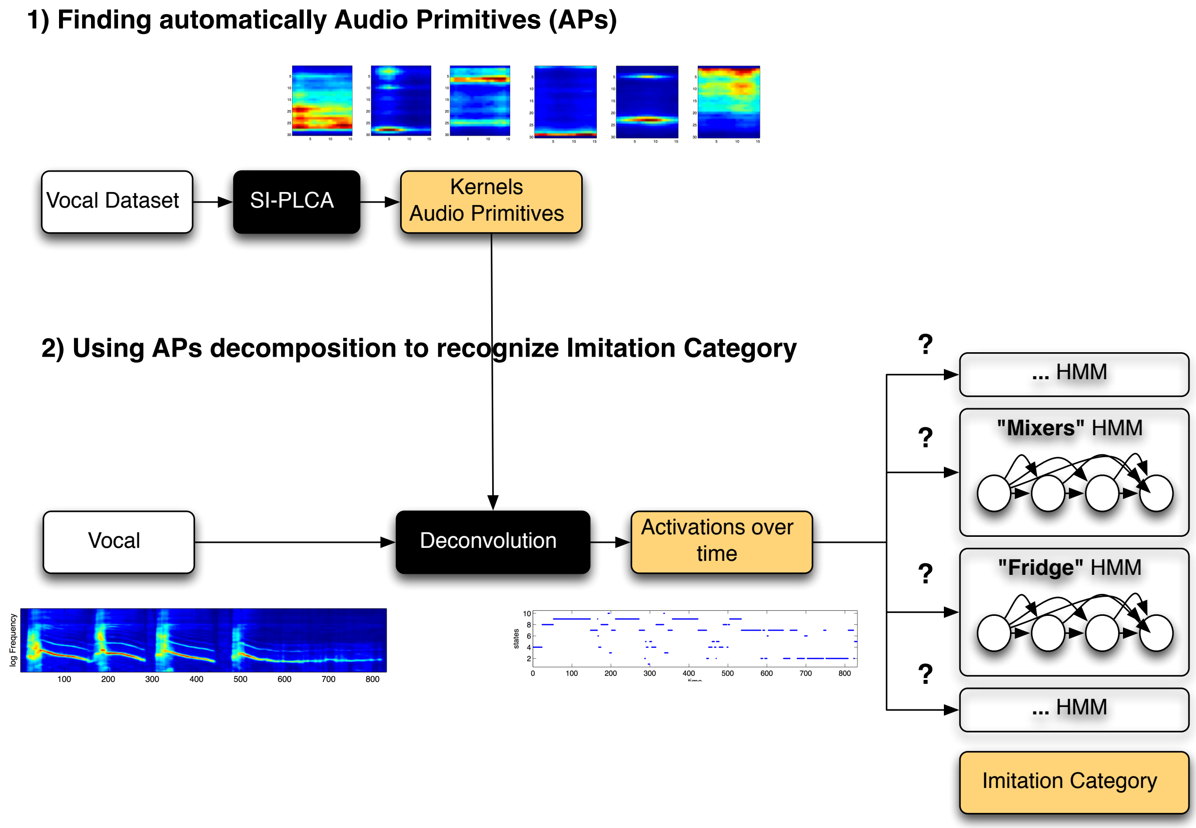

Figure 5: System for automatic extraction of audio primitives for HMM-based classification.

database, can be used to characterize efficiently the gestural primitives.

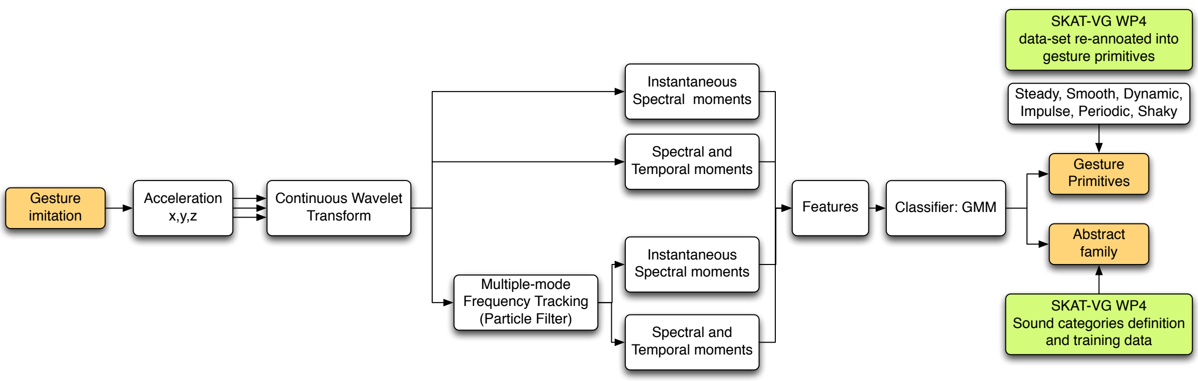

Figure 6: System for extracting gestural primitives from a gesture imitation.

1.3.5 Systems and software for sound sketching (WP6, WP7)

The SkAT-VG consortium has been developing sound synthesis tools that are suitable for simulating the sound sources represented by imitations, as well as high-level control strategies and layers to combine the sound models and to manipulate their parameters using vocalization and gestures.

The Sound Design Toolkit (SDT) is a library of physically informed sound synthesis models for education and research in sonic interaction design (Baldan et al., SoftwareX 6, 2017). The software can be considered as a virtual Foley pit to sketch interactive sonic experiences in gaming, product design, audio-visual production and many other contexts. Originally started and developed in the scope of EU projects SOb and CLOSED (2001 – 2003, 2006 – 2009), the SDT library has been extensively reworked in SkAT-VG, by enriching its sound palette, re-engineering the software architecture, redesigning the user interface, and adding sound analysis and post-processing capabilities to enable temporal control of the synthesizers through vocal and gestural imitations.

The SDT follows an ecological approach to everyday listening. The base assumption is that sounds carry meaningful information about the events responsible for their production, and that humans are able to interpret this information to gather knowledge about what happens around them (e.g. a car passing by, the locking of a door, or a storm approaching). Everyday sounds can be organized in a taxonomy, as in figure 7, that reflects the latest knowledge gathered in WP4 on the perception and categorization of everyday sounds and their imitations. The available sound models are grouped in four main classes of sounds following a criterion of causal similarity: vibrating solids, liquids, gasses, and machines. They also follow a hierarchical organization, where rich and evolving sound textures can be described as a combination of one or more lower level models. The synthesis algorithms make use of physically informed procedural audio techniques, which generate sound from a computational description of the sound producing event, rather than manipulating pre-recorded sound samples or wavetables. Physical models are functional to the ecological approach described above because they describe sound production in terms of its cause rather than its effect, focusing on the process rather than the product.

The SDT synthesizers are combined and parametrized into a collection of timbral families. A timbral family is defined as a set of one or more sound synthesis models, bound in its parameter space to represent a specific sound category. Sound categories have been experimentally assessed in WP4 as unambiguously discriminable sound events in terms of interaction, temporal and timbral properties. Timbral families can be expressively controlled by vocal and gestural imitations, by analyzing the input signals and extracting a set of timbral and motion descriptors which allow to act in real time on the synthesis parameters. Like the sythesis models, analysis tools are also organized in a hierarchical structure. Several low-level descriptors are combined to capture four higher level vocal primitives defined in WP3, namely the four main different ways used by humans to produce vocalizations: Phonation, myoelastic activity, air turbulence, and impulses.

SkAT Studio is a Cycling'74 Max modular framework, conceived at GENESIS to host diverse technologies developed by the partners of the project. The general workflow is composed of five stages, as described in Figure 8:

1. Input: vocal and gestural signals acquisition;
2. Analysis: audio and gestural feature extraction;
3. Mapping: transformation of the extracted features in control values for sound production;
4. Synthesis and processing: sound production;
5. Output: sound rendering and recording.

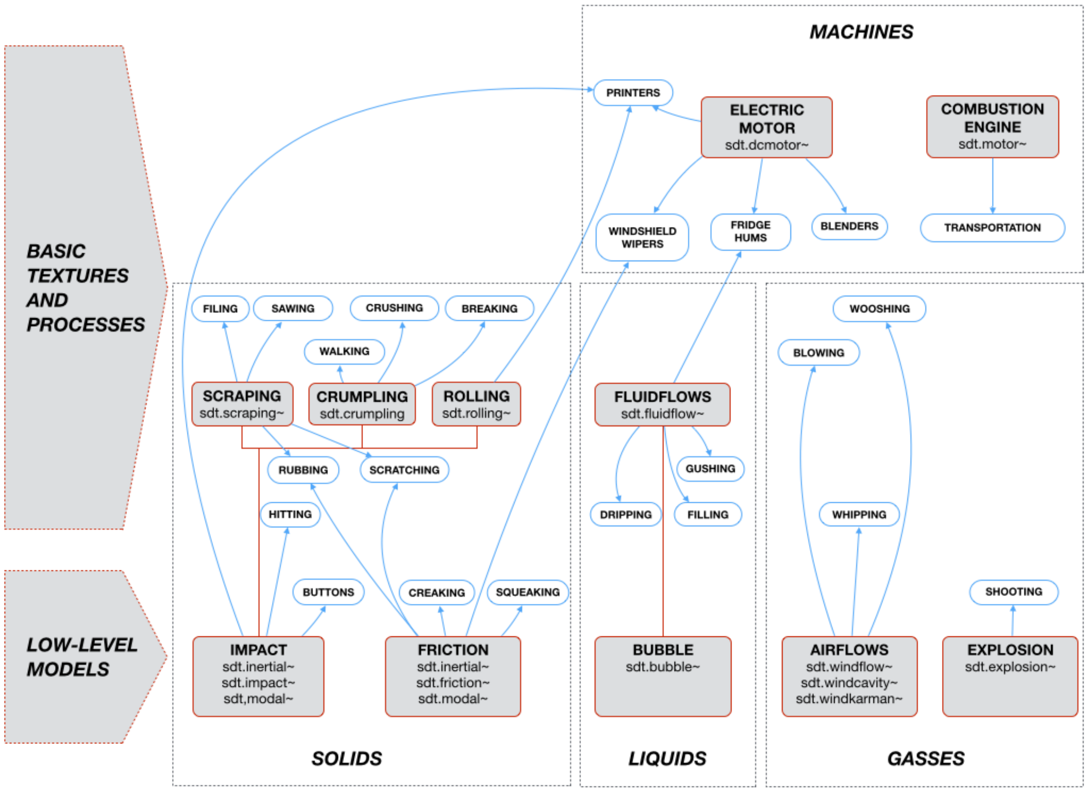

Figure 7: The SDT taxonomy of sound models. The bottom-up hierarchy represents the dependencies between low-level models and temporally-patterned textures and processes, for the four classes of sounds: Solids, liquids, gasses, and machines.

In the main GUI, shown in figure 9, 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). Signal and control data routing throughout the SkAT Studio framework can be programmed at will by means of patchbays. Modules (analysis, mapping, sound synthesis and processing) can be 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. Predefined control scenarios are based on synthesizers developed with the SDT toolbox.

miMic is a prototype application for vocal and gestural sonic sketching, designed to represent a sonic analogue of what paper and pencil are for visual sketching (Rocchesso et al., TEI, 2016). Incoming vocalizations are automatically classified into sound categories, which are used to select their corresponding timbral families. Once selected, timbral families can always be controlled in real-time by voice and gestures, allowing the sound designer to quickly explore a vast sonic space and intuitively produce expressive sonic sketches. As a successive step, sketches may be turned into more refined sound prototypes by further adjustment of the synthesis parameters through a dedicated graphical user interface. The application is composed of a hardware and a software part, and its design emerged from the observation and analysis of several workshops, design sessions and case studies. The process of sonic sketching could be roughly split into four stages:

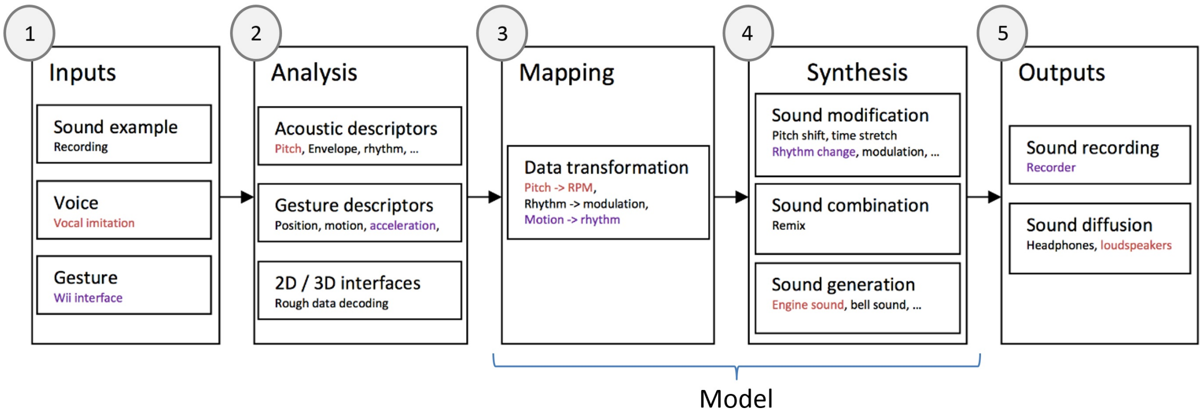

Figure 8: SkAT Studio 5-stages process. Two scenarios are highlighted in red and in purple. Red scenario: control of an engine sound using voice. Purple scenario: change of the rhythm of a sound using hand motion.

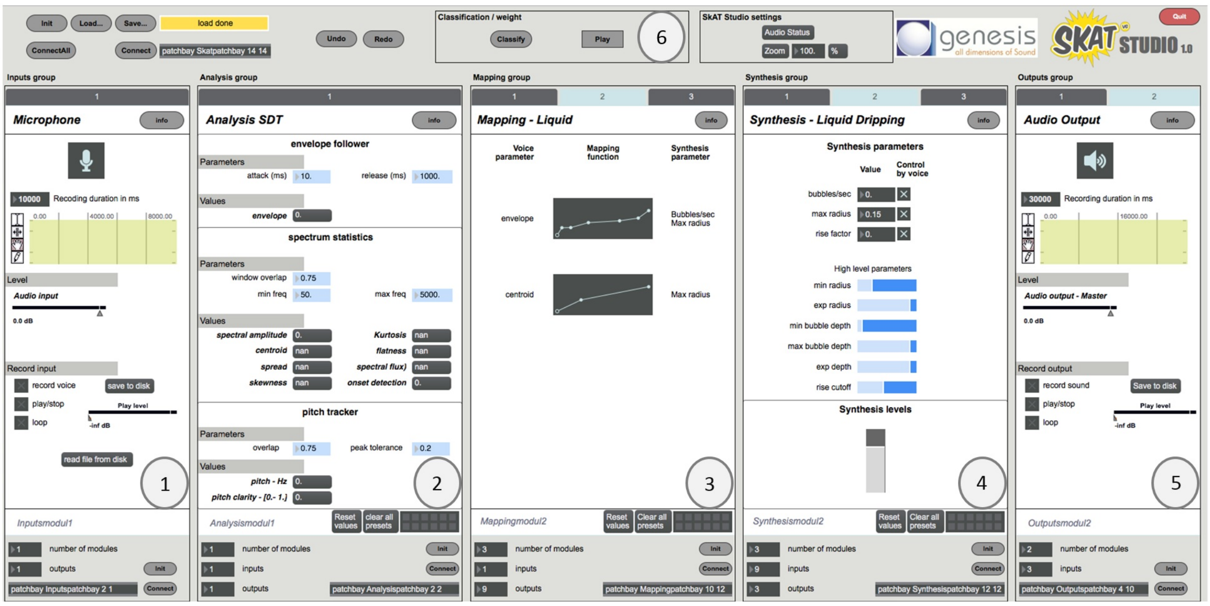

Figure 9: SkAT Studio GUI. The scenario is the control of a blowing model using voice.

1. Communication of a sound type by vocal examples and verbal specification;
2. Personification of the sound type, with the designer mimicking a sound behavior and the system mirroring it;
3. Expansion of the sonic space by exploration of new vocal control territories;
4. Saturation of the non-verbal conversation between designer and system. To refine the sketch, additional information need to be exchanged by means of conventional interface or scripting elements.

The software application provides two main modes of use: A *Select* mode, based on automatic classification, in which the user communicates (stage 1) a sound type to the system by vocal examples, and a *Play* mode in which the user expands (stages 2 and 3) and explores the sonic space offered by the system, based on the imitation-driven sound synthesis capabilities of the Sound Design Toolkit. The Select mode can be personalized by individual training of a GMM

Figure 10: Making a sonic sketch with miMic.

classifier, or one of the classifiers developed in WP5 can be used. In the *Play* mode, the vocal and gestural input is mapped to the available synthesis parameters through a control layer designed to reflect the research results of WP4, which show that humans are able to imitate only a few timbral parameters at the same time and that different timbral families require different imitation strategies, which in turn are best represented by different features and descriptors. For instance, vocal turbulence can be directly related to aerodynamic noises, impulsive behavior is often used to imitate impulsive sounds, like hitting or shooting, whereas phonation and precise frequency estimation is more relevant in pitched sounds, like those made by electric motors and combustion engines. At the end of the sketching process, a GUI offering separate and direct access to all the synthesis parameters covers the fourth and final stage, towards the creation of a prototype. The hardware part complements the software application, and it consists of an augmented microphone with embedded inertial sensors and buttons. The buttons placed on top of the microphone allow to select the two different operation modes, eliminating the need of a visual display until the final refinement stage. At the same time the embedded intertial measurement unit captures acceleration and orientation signals, to detect the ancillary gestures which often accompany vocal imitations and are normally used to reinforce and complement vocalizations. Acceleration and orientation signals can be directly coupled to sound model parameters, or combined into higher level gestural features. Some of the most relevant qualities include energy, smoothness and timing of major strokes.

MIMES is another prototype application for sonic sketching that makes use of gestures to control in real time the manipulation and resynthesis of pre-recorded sounds, including vocal imitations. The system is composed of a software part developed in Cycling'74 Max and a hardware tangible interface composed of sensorized, 3d-printed objects. The sound generation system relies on corpus-based concatenative synthesis rather than physically informed models. The application workflow of MIMES is similar to the one found in miMic, as it consists of two main modes of use: A *Record* step in which the user records a vocalization to propose an initial sound morphology, and a *Control* step in which the user can produce a sonic sketch by manipulating the tangible objects. Another similarity with miMic is the presence of a GUI which exposes all the relevant synthesis parameters and temporal controls for later manual refinement of the sonic sketch. In the Record mode a vocal imitation is recorded in a buffer,

Figure 11: Making a sonic sketch with MIMES.

segmented into small fragments and analyzed to obtain a time series of audio descriptors. The sound descriptors are then sampled and fed to the sound synthesis engine at a fixed rate. The synthesis engine is responsible for the generation of sounds reflecting the timbral properties of the vocal imitation. MIMES implements a corpus-based concatenative sound synthesis engine, exploiting a technique which is similar to audio mosaicing. This sound synthesis engine relies on specific sound databases, previously segmented and analyzed, referred to as corpora. For each frame of sound descriptor values, the synthesis engine plays a small sound fragment belonging to one of its corpora, selected to match as closely as possible the timbral features of the input. It is worth noting that the recorded vocal imitation can be considered itself as a corpus, and therefore directly used as source material for the concatenative resynthesis process. In Control mode, the sensorized tangible objects are used to modify the sonic sketches. First, movement analysis is performed to obtain higher level gesture descriptors from the raw input coming from the sensors, such as the orientation or the movement intensity. These movement descriptors can alter the original sound descriptor profiles or directly control the synthesis parameters, namely corpus weights and sound descriptor weights. More complex control actions can be exerted by using the mapping-by-demonstration method. In this case, an example gesture is recorded and mapped to a vocal imitation by a HMM. This relation can then be used to regenerate the sound descriptors when replaying the gesture. Performing the gesture with some variations will generate variations in the sound descriptors as well, and consequently in the final sound sketch. The temporal evolution of gestural control can be recorded and played back as well. The generated descriptors can be individually looped or manipulated in real time, preserving the desired timbral properties of the resulting sound and tweaking the others, thus allowing an iterative refinement of the sonic sketch.

1.3.6 Sound design processes (WP7)

WP7 ("Sonic Interaction Design") is the Work Package that has been studying the use and exploitation of vocal utterances and gestures from the design viewpoint. It represents the playground, shared among the other WPs, in which basic research and scientific findings have been distilled in relevant and systematized vocal and gestural sketching practices. New scientific knowledge and software modules have been used to inform design research and to assess the emerging sound design tools and methods. WP7 has been playing the role of a hub, a dispatcher, by providing design feedbacks that are reflected in the experiments and studies of other work packages, thus establishing a shared, methodological connection between scientific and design research.

The main achievements of design research in WP7 are (D7.7.2):

A propaedeutics on vocal sketching: a reference framework of design exercises and tools, grounded in phonetics and auditory perception of vocal imitations;

The unfolding and understanding of the creative process of vocal sketching: a set of interviews with expert sound practitioners, and two main studies on the cognitive behavior of novice and expert sound designers, engaged in problem-solving a specific sound design task;

Collaborative and embodied sketching tools: Cooperation, shared creativity and embodied control are the main ingredients of demonstrators, software prototypes, and side applications developed in SkAT-VG.

Workshops have been the major means to develop and assess the SkAT-VG methods and tools (Delle Monache et al., ICAD, 2015; Houix et al., Audio Mostly, 2016). Several concrete sound design experiences have been performed, and a growing community of novice and expert sound designers have had the opportunity to experiment the methods and tools, and provide valuable insights and feedback. For example, Figure 12 shows a collaborative design session that was held in Paris, as part of an industrial case study with the French car manufacturer PSA.

The emerging SkAT-VG methods have been systematized in design exercises arranged in a structured workshop format. From the analytical introduction to the use of voice for sound communication, to the vocal imitation game (see Figure 13), to rather design-oriented assignments, the SkAT-VG workshop unfolds a path of incremental embodied and sharable experiences around creativity in sound design. The practice of vocal imitations is exploited as a sensory-motor instrument to access the internal representations of one's own auditory experiences and make sonic intentions audible.

Protocol analysis Protocol and linkographic analysis of the efficiency of the creative processes have been applied for the first time to the sound design domain (Delle Monache and Rocchesso, Audio Mostly, 2016), and exploited to assess the most compelling sessions. These tools provided relevant qualitative and quantitative information on the embodied and cooperative aspects of the vocal sketching methodology and on the effectiveness of the SkAT-VG prototypes as mediation tools to support sound creativity. It emerged that novice sound designers are more prone to use vocalizations in solving sound design tasks, yet they lack of a proper attitude towards iterative evaluation. The process of expert sound designers, instead, reflects the current state of the sound design workflow, essentially based on verbal specifications and a dense recourse to evaluation to face pressing time-constraints in an activity, sound

Figure 12: A collaborative design session for driving-aid sounds in cars

production, which is rather time-consuming. In both cases, the sharable, embodied and immediate aspects of vocal sketching fostered the building of effective cooperative practices in the definition of sound specifications. Eventually, the progressive understanding of the unfolding of sound design-thinking in early design stages provided new knowledge that found application in the development and evaluation of the sonic embodied tools, from SkAT Studio, miMic and MIMES to SEeD.

Figure 13: The vocal imitation game

Embodied Sound Design is a process of sound creation that extensively involves the designer's body. In particular, voice and gestures are used to rapidly produce sonic sketches that other humans can interpret, but that can also be transformed to synthetic sounds. In order to explore how vocal utterances may be automatically converted to synthetic sounds, IUAV conceived an artistic installation, called *S'i' fosse suono*, where sixteen brief vocal self-portraits are arranged in the form of an audiovisual checkerboard, depicted in figure 14. The recorded non-verbal vocal sounds were used as sketches for synthetic renderings, realized by the sound design professional Andrea Cera using physics-based modeling (SDT) and corpus-based synthesis as reference sound modeling techniques. The constraints given to the sound designer in terms of usable sound models, and his use of some automatic feature extractors, made the eventual automation of the process easier to foresee. The art installation was conceived to envision how vocal sketching may be used to design sounds with a vast timbral palette, as it is given by versatile sound synthesis models. Moreover, it was possible to experimentally assess the versatility of the two voice-driven sound synthesis methods to render an ample spectrum of sounds, from abstract to concrete (D7.7.2). This sound design process was then automated to

Figure 14: *S'i' fosse suono* with two sliders superimposed for evaluating concreteness and naturalness

create SEeD, a tool that allows the designer to first set physics-based or corpus-based sound models, and then to play interactively with them to create elaborate sounds (D7.7.2). SEeD, whose construction relies on the experiences of miMic and MIMES, is schematically described in figure 15. A preliminary release of SEeD was used in the 48 hours of sound design in Château La Coste. The complementary character of the two synthesis methods emerged from the observed workflow and from interviews, and important suggestions for evolving SEeD into a key exploitable result were collected.

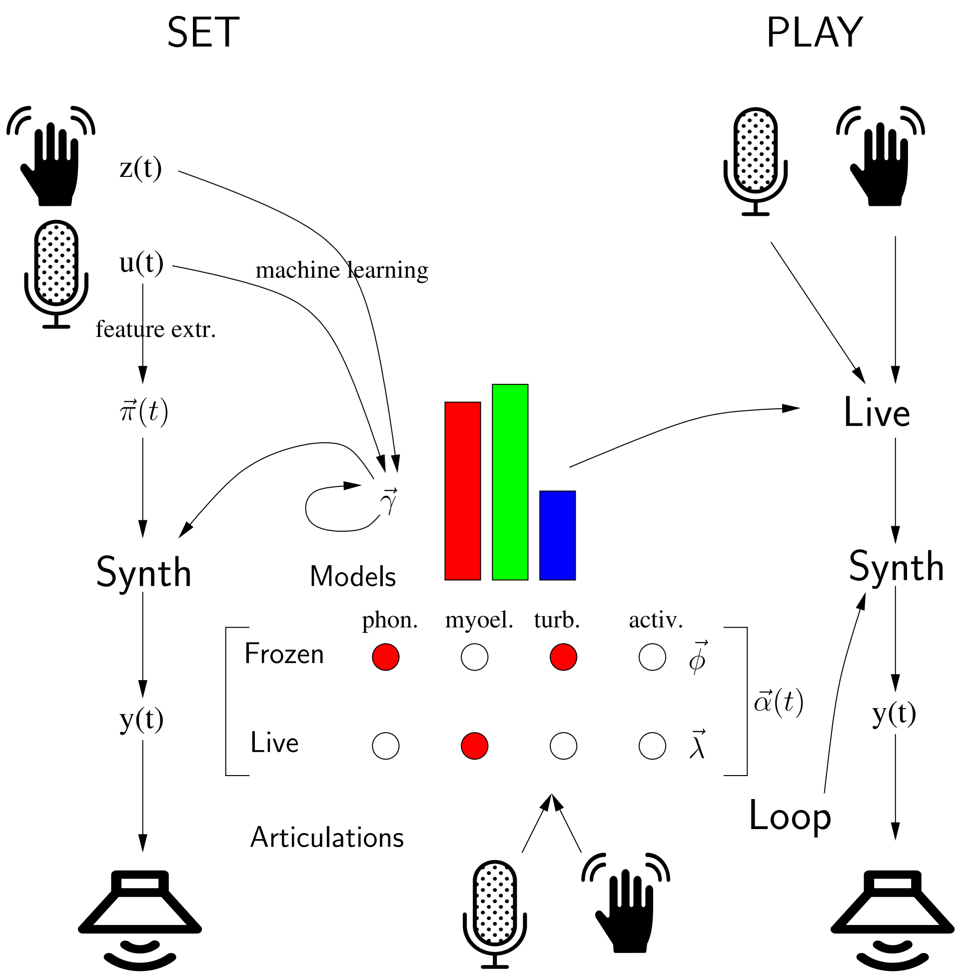

Figure 15: Sketch of SEeD

1.4 Potential impact

The SkAT-VG project has a long-term vision: To introduce, facilitate, and spread the use of non-verbal vocalizations and expressive manual gestures in sound design workflows, especially in the early sketching stages and in cooperative processes. The scope of sound design is very broad, including any context where the sonic behavior of objects is relevant for their use and aesthetics. In this respect, a sonic interactive object could be anything from a videogame to a car. The vision of SkAT-VG was that technologies that enable fast and intuitive prototyping, refinement, and evaluation of product sounds would greatly facilitate the process of industrial design, boost the creativity of sound designers, and improve the quality of our sound environment. At the end of SkAT-VG, a bag of tools, design knowledge, and practices have been introduced that would allow the designer to use vocal and manual gestures to create synthetic, model-based sounds. A network of professionals has been exposed to SkAT-VG methods and tools, and workshops such as the 48 hours of sound design in Château La Coste or the industrial case study at PSA created a mutual binding between SkAT-VG and sound professionals. New practices are likely to emerge out of this process of convergence between science and creative practice.

One important goal of the SkAT-VG project was to foster collaborative work in teams of professionals reflecting on sound problematics. The initial observation was that most of the sound design activities involve a process of personal and internal creation, despite the fact that ensembles of people are very good at cooperating to create beautiful sonic outcomes (e.g., choirs). By bringing vocal and gestural sketching tools that are easy to grasp, SkAT-VG aims at making a whole team working together since the early beginning of a project, to create sonic behaviors that would be impossible to prototype by a person alone. In creative workflows, interactions between different actors are valuable to open new voices that are not accessible when working alone. Spreading these cooperative sound design practices, however, requires further dissemination work and interaction with educational institutions. Some of the SkAT-VG researchers are also active educators, and will continue playing the main role of evangelists of sound sketching by voice and gesture.

If SkAT-VG is followed by a sustained flow of research and dissemination, and if a number of early adopters contribute to its diffusion, the project will determine significant advances in Europe in the design practices for a variety of products, such as films and multimedia shows (sound effects), games (sound-mediated sense of agency), everyday products (sonic affordances and aesthetics), environments (soundscapes), human-machine interfaces, and vehicles. The concepts developed during the project are also believed to be applicable to areas other than sonic sketching. In fact, as SkAT-VG has developed tools to infer user's intentions from voice and gestures, it will be possible to consider vocal imitations and manual gestures more generally for expressive and intuitive human-computer interactions (including new interfaces for musical instruments and creative interfaces for non-professionals) in different fields of application. Moreover, as it often happens when the expressive abilities of humans are exploited through technology, new unforeseeable applications and activities naturally emerge. For example, during the course of the project, possible routes of exploitation that emerged from research and workshops are board games (a guessing game analogue to Pictionary has been fully developed and tested in SkAT-VG, see Figure 13) and aids for language teaching (the SkAT-VG experiment at Språkstudion of Stockholm University shed light on possible uses of vocal imitations to teach foreign languages).

The impact of SkAT-VG on science is likely to develop over a few years after project conclusion. Vocal imitations have been emerging as a probe to study auditory perception at large, and are now a key ingredient of experiments that support current theories of embodied cognition and human language development. If we search for embodied representations of sounds, we cannot avoid considering the human voice as the action component of perception-action ensembles. The most salient sound-producing physical phenomena (friction, turbulence, impact, etc.) find counterparts in voice production mechanisms, which act as representational anchors of the outside sonic world. Through vocal imitations and auxiliary gestures, two persons can communicate and share auditory experiences. Vocalizations and gestures open a window into sound cognition, where sensory and memory representations of sounds interact through action. Some research threads along this line have been initiated in the SkAT-VG project and are likely to develop further in the years to come.

Often sound designers are sound artists as well. They wear different hats when creating the 'voice' of an electric car or when conceving an audiovisual installation, but they are dealing with the same materials and are using much of the same tools. Both activities are expressions of human creativity, although one is ruled by constraints and iterative evaluation, while the other is free to use whatever serves the sake of a vision. Indeed, artists' visions are useful to foresee possible new behaviors, practices, and products. The SkAT-VG project deliberately exploited artistic visions when developing its embodied sound design process and tools, and collaboration between scientists and artists is likely to develop further, encouraged by physical contiguity in research and culture institutions, such as those of the SkAT-VG Consortium.

The Final Event of the SkAT-VG project is the Sound Design Rendezvous in Paris, on january 19, 2017. The program of the day, reported in table 1 gives a glimpse on what the project has achieved and, more importantly, on the trails it is going to leave in science, design, innovation, and art. Videos of the presentations are available online: `http://medias.ircam.fr/x7f9fa7`.

Thursday, January 19, 2017 - IRCAM, Paris The Sound Design Rendezvous		
09.30-10.00	Introduction	Davide Rocchesso (SkAT-VG coordinator)
10.00-11.30	**What we learned**	
	"Production of vocal imitations"	Sten Ternström (KTH)
	"Perception of vocal and gestural imitations"	Guillaume Lemaitre (IRCAM)
	"Classification of vocal and gestural imitations"	Geoffroy Peeters (IRCAM)
	"Sound design"	Stefano Delle Monache (IUAV)
11.30-12.15	"Sound design for everyday interactions"	Elif Özcan (Delft Univ. of Technology)
12.15-14.00	**What we produced**	
	Demonstrations	Stefano Baldan (IUAV) and Clément Dendievel (GENESIS)
14.00-14.45	"Expressive sound design for music-based biofeedback"	Marc Leman (Ghent University)
14.45-15.00	48 hours of Sound Design at Château La Coste	short documentary by Sylvestre Miget
	The Brown Lipstick Sketches	audiovisual piece by Andrea Cera
15.00-16.00	**Where we are heading to**	
	Research	Olivier Houix (IRCAM)
	Practice	Frederic Bevilacqua (IRCAM)
	Innovation	Patrick Boussard (GENESIS)
16.00-17.00	Practices of Sound Design	Round Table
	Industry	William Rodriguez (Renault)
	Creation	Andrea Cera (sound designer and composer)
	Pedagogy	Roland Cahen (ENSCI – Les Ateliers)

Table 1: Program of the Sound Design Rendezvous in Paris, 2017

1.5 Contacts and website

Primary contact: Davide Rocchesso (Project Coordinator, Iuav University of Venice, Italy): roc@iuav.it

Project Consortium:

Iuav University of Venice, Italy

Institut de Recherche et Coordination Acoustique/Musique, Paris, France

KTH Royal Institute of Technology, Stockholm, Sweden

Genesis, Aix en Provence, France

Project webpage: www.skatvg.eu for contacts, demos, advances, news, events, journal and conference publications, deliverables, updates, etc.

Twitter: https://twitter.com/skatvg for @SkATVG project-related events.

2 Use and dissemination of foreground

Even though no specific Work Package has been exclusively designed for it, dissemination of project findings and results has been considered to be crucial to success of the SkAT-VG project. Every partner has had a portion of the budget dedicated to cover the expenses for publications and promotion. Rules for dissemination and use, such as confidentiality, mandatory communications, and ownership of results are defined in the Consortium Agreement.

2.1 Section A

2.1.1 Dissemination strategy: Objectives and goals; Tools and actions

Website
The general project strategy for dissemination is based on openness: project reports, experimental data, models, and tools have been made available and will continue to be made available to all through the project website `http://www.skatvg.eu`. This approach will enable independent evaluation and cross-validation of results. Moreover, the availability of research tools may convince some practitioners to embark on new approaches to sonic interaction design based on vocal sketching and gestural interaction. The SkAT-VG website will continue to be the main source of dissemination of research results, demonstrations, and software tools.

Seminars and training
Since the early stages of SkAT-VG there has been an extensive involvement of external stakeholders, such as expert sound designers, whose advice has been valuable to better tailor the design of sketching tools. Another important class of stakeholders that have been involved since the beginning of the project is that of expert vocal imitators and professionals who make extensive use of vocal abilities. They have been indispensable for the identification of the most relevant vocal primitives. Sound and interaction designers, both students and professionals, have been involved in workshops where vocal and manual sketching has been at the center of design exercises. The exercises have been directed toward the definition of scenarios and applications. Then, actual designs of sonic interactive objects have been performed and documented, with involvement of the Sonic Interaction Design community.

Interactions with other projects, networks and technology platforms
All the reported work in voice production, perception, and machine learning has been done in SkAT-VG, with no interaction with other sources of funding. The work on gestures has benefited from previous and ongoing work at IRCAM, such as the use of motion sensor to control sound. Nevertheless, the studies on vocal and gesture imitation and the analysis of the database have been completely done within SkAT-VG. In particular, a totally new approach was developed to analyse gesture in real-time based on wavelets. The analysis of the different multimodal strategies the listeners use to describe sound was performed entirely within SkAT-VG. The classification work has been largely based on the SkAT-VG studies on production and perception of vocal imitations, and brand new features have been derived for the project purposes. The SkAT-VG approach to classification now models the time evolution of sounds, thus providing a constructive solution to an open scientific problem. The automatic prediction of articulatory classes, as developed at KTH, is based on the auditory receptive fields toolbox developed in a previous project. However, all articulatory models of the phonetic, myoelastic, and turbulent components including all used audio features are new and have been developed in Matlab in SkAT-VG, with a realization adapted to real time at IUAV. The work on sound models at IUAV is the continuation of over a decade of studies and developments on sound synthesis by physical modeling, partly funded by previous FET and NEST projects (SOb and

CLOSED). The Sound Design Toolkit, as it has been distributed by SkAT-VG in 2016, is a complete redesign that contains several new sound models that are necessary to represent the sound categories emerging from studies in perception and production of vocal imitations, as well as brand new feature extractors. The Sound Design Toolkit distributed by SkAT-VG is composed of a core framework entirely developed in ANSI C and a collection of wrappers for Max and PureData. The code is designed to be portable across different operating systems (Windows, Mac OS X, Linux), and the APIs exposed by the core framework allow the reuse of the synthesis algorithms in a wide variety of developing environments other than Max and Pd. The software SkAT-Studio has been totally developed within SkAT-VG, as well as the new LEA Sound Design Module based on drawing in the time-frequency plane. Both software projects, whose development has been led by GENESIS, were largely based on the analysis of the interviews that were collected from sound designers in the SkAT-VG project. The prototypes miMic, Mimes, and SEeD have been completely developed within SkAT-VG.

SkAT-VG has been developing knowledge, methods and tools that are partly derived from previous researches carried out by the Consortium. In particular, Table 2 summarizes the exploitation of technologies and data previously developed or collected in other projects. All the products, tools, and datasets used in SkAT-VG, mentioned in this Periodic Report and not mentioned in Table 2, have been entirely developed or collected in SkAT-VG.

Dissemination of results

During its lifetime, the SkAT-VG project has been accumulating scientific knowledge in different fields, which have been thoroughly documented in the project deliverables. Some of the findings have been disseminated through scientific conference and journal papers, which are listed in Table A1 of section 2.1.2. Some research results are in press or being submitted for publication, and support for open access will be seeked through the OpenAIRE FP7 Post-Grant Pilot. Some other results require further elaboration before being disseminated through scientific publications. In particular, the work on machine learning and production of vocal imitations is likely to produce relevant scientific articles. The availability of a large multi-signal database of vocal and gestural imitations will spur further research in this area and will become an important reference for studies in sound perception. In perspective, it is expected that the scientific trail of SkAT-VG will last for several months after its conclusion.

The diffusion of sound sketching practices among practitioners requires a continuing effort for organizing educational activities in design schools and workshops with professionals. This will be made more effective by the reduction of SkAT-VG tools and methods to a well-tested core. During the project, several software modules, prototypes, and demonstrations have been developed, but most of them are to be considered as intermediate steps toward the creation of practical design instruments. After the project conclusion, the sound-sketching workshop format will be formally defined and accompanied by a fully-produced board game. The prototype SEeD will be refined and an exploitation strategy will be activated to transform it into a product.

Partner	Object:	Related Project:
IUAV	Sound Design Toolkit	Started being developed in project IST-2000-25287 (The Sounding Object). Further developed in project FP6-NEST-PATH-29085 (CLOSED). SkAT-VG has been providing an extension of the palette of models and a re-writing of most models: at least 50% extension of prior work.
IRCAM	Collection of everyday sound categories	Defined in projects FP6-NEST-PATH-29085 (CLOSED) and Sample Orchestrator (ANR France). Exploited in SkAT-VG as follows: Selecting a subset of the categories, populating these categories with exemplars, conducting identification experiment to select only the categories that are not confused and the best exemplars within each category. About 90% of the work is new in SkAT-VG.
IRCAM	Physical objects used in Mimes	The physical objects for the Mimes installation/demonstration were designed during a previous project (ANR Legos) to study sensori-motor learning. These interactive objects have been further developed, adding the voice component, and have been used in a completely different research context, with very different goals.
IRCAM	MuBu	The MuBu multi-buffer is a container for sound and motion data. It provides a structured memory for the real-time processing of recorded sound and action together with interfaces and operators as a set of complementary Max externals. The ensemble of MuBu externals for Max allows for sound synthesis such as granular, concatenative and additive synthesis and interactive machine learning. In SkAT-VG, some externals were added to perform real-time wavelet analysis of gestures, and improved interactive machine learning.
KTH	ELAN annotation procedures	ELAN is provided by The Language Archive project at the Max Planck Institute for Psycholinguistics, Nijmegen, NL. The annotation procedures have been developed in SkAT-VG.
Genesis	XTract	Largely developed within the projects FET-Open-255931 (UNLocX) and EU-FP7-233980 (BESST). SkAT-VG provided a 10% extension and allowed industrialization of the product.
Genesis	Active Sound Design	Largely developed on Genesis own funds. SkAT-VG allowed to improve the overall measure process.
Genesis	LEA ("the Sound Lab for Industry")	A new module for LEA has been developed from the indications of the SkAT-VG interviews with sound designers.

Table 2: Relations with other projects.

2.1.2 List of publications

TABLE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS , STARTING WITH THE MOST IMPORTANT ONES										
NO.	**Title**	**Main author**	**Title of the Periodical or the Series**	**Number, date or frequency**	**Publisher**	**Place of publication**	**Year of publication**	**Relevant pages**	**Permanent identifiers (if available)**	**Is/Will open access provided to this publication?**
1	The Sound Design Toolkit	S. Baldan, S. Delle Monache and D. Rocchesso	Software X	In press	Elsevier		2017			Yes
2	Vocal imitations of non-vocal sounds	G. Lemaitre, O. Houix, F. Voisin, N. Misdariis, and P. Susini	PLoS ONE	11:12	PLoS	San Francisco, CA, USA	2016	e0168167	http://dx.doi.org/10.1371/journal.pone.0168167	Yes
3	Non-speech voice for sonic interaction: a catalogue	A. Del Piccolo and D. Rocchesso	Journal on Multimodal User Interfaces		Springer	New York, NY, USA	2016	1–17	http://link.springer.com/article/10.1007/s12193-016-0227-6	Yes
4	Organizing a sonic space through vocal imitations	D. Rocchesso, D.A. Mauro and C. Drioli	Journal of the Audio Engineering Society	64:7/8	AES Press	New York, NY, USA	2016	474–483	http://www.aes.org/e-lib/browse.cfm?elib=18333	Yes
5	Vocal imitations of basic auditory features	G. Lemaitre, A. Jabbari, N. Misdariis, O. Houix, and P. Susini	The Journal of the Acoustical Society of America	139(1)	AIP Publishing LLC	Melville, NY, USA	2016	290–300	http://dx.doi.org/10.1121/1.4939738	No
6	Multisensory texture exploration at the tip of the pen	D. Rocchesso, S. Delle Monache and S. Papetti	International Journal of Human-Computer Studies	85	Elsevier	Amsterdam, Netherlands	2016	47–56	http://dx.doi.org/10.1016/j.ijhcs.2015.07.005	Yes
7	Sketching sound with voice and gesture	D. Rocchesso, G. Lemaitre, P. Susini, S. Ternström and P. Boussard	Interactions	22:1	ACM	New York, NY, USA	2015	38–41	http://dl.acm.org/citation.cfm?id=2685501	Yes
8	Idealized Computational Models for Auditory Receptive Fields	T. Lindeberg and A. Friberg	PLoS ONE	10:3	PLoS	San Francisco, CA, USA	2015	e0119032	http://dx.doi.org/10.1371/journal.pone.0119032	Yes
9	Bauhaus legacy in Research through Design: the Case of Basic Sonic Interaction Design	S. Delle Monache and D. Rocchesso	International Journal of Design	8:3		Taipei City, Taiwan	2014	139–154	http://ijdesign.org/ojs/index.php/IJDesign/article/view/1543	Yes
10	Sounding objects in Europe	D. Rocchesso	The New Soundtrack	4:2	Edinburgh University Press	Edinburgh, Scotland, UK	2014	157–164	http://dx.doi.org/10.3366/sound.2014.0060	No
11	On the effectiveness of vocal imitations and verbal descriptions of sounds	G. Lemaitre and D. Rocchesso	The Journal of the Acoustical Society of America	135:2	AIP Publishing LLC	Melville, NY, USA	2014	862–873	http://dx.doi.org/10.1121/1.4861245	No
12	Might as Well Jump: Sound Affects Muscle Activation in Skateboarding'	P. Cesari, I. Camponogara, S. Papetti, D. Rocchesso and F. Fontana	PLoS ONE	9:3	PLoS	San Francisco, CA, USA	2014	e90156	http://dx.doi.org/10.1371/journal.pone.0090156	Yes
13	Auditory bubbles reveal sparse time-frequency cues subserving identification of musical voices and instruments	V. Isnard, C. Suied, and G. Lemaitre	The Journal of the Acoustical Society of America. Proc. ASA Meeting.	140:4	AIP Publishing LLC	Melville, NY, USA	2016	3267	http://dx.doi.org/10.1121/1.4970361	No
14	Comparing identification of vocal imitations and computational sketches of everyday sounds	G. Lemaitre, O. Houix, F. Voisin, N. Misdariis, and P. Susini	The Journal of the Acoustical Society of America. Proc. ASA Meeting.	140:4	AIP Publishing LLC	Melville, NY, USA	2016	3390	http://dx.doi.org/10.1121/1.4970854	No
15	Cooperative sound design: A protocol analysis	S. Delle Monache and D. Rocchesso	Proceedings of the Audio Mostly	Norrköping, Sweden, October 04–06	ACM	New York, NY, USA	2016	154–161	http://dl.acm.org/citation.cfm?doid=2986416.2986424	No
16	Innovative Tools for Sound Sketching Combining Vocalizations and Gestures	O. Houix, S. Delle Monache, H. Lachambre, F. Bevilacqua, D. Rocchesso, and G. Lemaitre	Proc. of the Audio Mostly	Norrköping, Sweden, October 04–06	ACM	New York, NY, USA	2016	12–19	http://dl.acm.org/citation.cfm?doid=2986416.2986442	No

17	Sonic in(tro)spection by vocal sketching	A. Cera, D. Andrea Mauro, and D. Rocchesso	Proc. of the XXI Colloquio di Informatica Musicale	Cagliari, Italy, September 28 - October 01	AIMI	Venice, Italy	2016			Yes
18	Understanding cooperative sound design through linkographic analysis	S. Delle Monache and D. Rocchesso	Proc. of the XXI Colloquio di Informatica Musicale	Cagliari, Italy, September 28 - October 01	AIMI	Venice, Italy	2016			Yes
19	Sketching Sonic Interactions by Imitation-Driven Sound Synthesis	S. Baldan, S. Delle Monache, D. Rocchesso and H. Lachambre	Proceedings of the Sound and Music Computing conference (SMC)	Hamburg, Germany, August 31 - September 3			2016		http://smcnetwork.org/node/1994	Yes
20	Sketching Step in Sound Design: the Sound Designers' Point of View	P. Boussard, C. Dendievel and H. Lachambre	Proceedings of Internoise	Hamburg, Germany, August 21–24	Institute of Noise Control Engineering	Reston, VA, USA	2016	4505–4513	http://pub.dega-akustik.de/IN2016/data/articles/000187.pdf	Yes
21	Embodying sounds: Building and analysis of a database of gestural and vocal imitations	H. Scurto, G. Lemaitre, J. Françoise, F. Bevilacqua, P. Susini, and F. Voisin	Proceedings of the 7th Conference of the International Society for Gesture Studies (ISGS)	July 18–22, Paris	University of Texas at Austin	Austin, TX, USA	2016	269		No
22	A database of articulatory annotations of vocal imitations	P. Helgason, G. Laís Salomão, and S. Ternström	Proceedings of the XXIX Swedish Phonetics Conference (FONETIK)	Stockholm, Sweden, June 13–15			2016		http://fonetik.se/papers/f2016-6.pdf	Yes
23	miMic: The microphone as a pencil	D. Rocchesso, D. Mauro, and S. Delle Monache	Proceedings of the10th International Conference on Tangible, Embedded and Embodied Interaction (TEI)	Eindhoven, The Netherlands, February 14–17	ACM	New York, NY, USA	2016	357–364	http://dl.acm.org/citation.cfm?id=2839467	Yes
24	A set of audio features for the morphological description of vocal imitations	E. Marchetto and G. Peeters	Proceedings of the 18th International Conference on Digital Audio Effects (DAFx)	Trondheim, Norway, November 30 - December 3			2015	207–214	http://www.ntnu.edu/documents/1001201110/1266017954/DAFx-15_submission_64.pdf	Yes
25	Reverberation still in business: Thickening and propagating micro-textures in physics-based sound modeling	D. Rocchesso, S. Baldan, and S. Delle Monache	Proceedings of the 18th International Conference on Digital Audio Effects (DAFx)	Trondheim, Norway, November 30 - December 3			2015	12–18	http://www.ntnu.edu/documents/1001201110/1266017954/DAFx-15_submission_24.pdf	Yes
26	Vocal imitations of basic auditory features	G. Lemaitre, A. Jabbari, O.Houix, N. Misdariis, P. Susini	The Journal of the Acoustical Society of America. Proc. ASA Meeting.	137:4	AIP Publishing LLC	Melville, NY, USA	2015	2268	http://dx.doi.org/10.1121/1.4920282	No
27	Combining gestures and vocalizations to imitate sounds	H. Scurto, G. Lemaitre, J. Françoise, P. Susini, F. Bevilacqua	The Journal of the Acoustical Society of America	138:3	AIP Publishing LLC	Melville, NY, USA	2015	1780	http://dx.doi.org/10.1121/1.4933639	No
28	Analyzing and organizing the sonic space of vocal imitation	D.A. Mauro and D. Rocchesso	Proceedings of the Audio Mostly	Thessaloniki, Greece, October 07–09	ACM	New York, NY, USA	2015	23	http://dl.acm.org/citation.cfm?id=2814921	No
29	To "Sketch a Scratch"	A. Del Piccolo, S. Delle Monache, D. Rocchesso, S. Papetti and D.A. Mauro	Proceedings of the 12th Sound and Music Computing Conference (SMC)	Maynooth, Ireland, July 26 - August 01			2015		http://smcnetwork.org/node/1889	Yes
30	Growing the practice of vocal sketching	S. Delle Monache, D. Rocchesso, S. Baldan, D.A. Mauro	Proceedings of the 21st International Conference on Auditory Display (ICAD)	Graz, Austria, July 07–10	Georgia Institute of Technology	Atlanta, GA, USA	2015	58–65	http://hdl.handle.net/1853/54105	Yes
31	Advanced signal processing methods for the analysis of transient radiated noise from submarines	T. Leissing, C. Audoly, H. Lachambre and G. Stempfel	Proceedings of InterNoise	Melbourne, Australia, November 16–19	Institute of Noise Control Engineering	Reston, VA, USA	2014	852–861	http://www.acoustics.asn.au/conference_proceedings/INTERNOISE2014/papers/p113.pdf	Yes

32	Physically informed car engine sound synthesis for virtual and augmented environments	S. Baldan, H. Lachambre, S. Delle Monache, and P. Boussard	Proc. of the 2nd Workshop on Sonic Interactions in Virtual Environments - IEEE VR 2015	Arles, France			2015			No
33	Non-Verbal Imitations as a Sketching Tool for Sound Design	G. Lemaitre, P. Susini, D. Rocchesso, C. Lambourg, P. Boussard	Lecture Notes in Computer Sciences	8905	Springer	Berlin, Germany	2014	558–574	http://link.springer.com/chapter/10.1007/978-3-319-12976-1_34	No
34	Self-organizing the space of vocal imitations	D. Rocchesso and D.A. Mauro	XX Colloquio di Informatica Musicale	Rome, Italy, October 20–22	AIMI	Venice, Italy	2014	124–127	http://smc.dei.unipd.it/cim_proceedings/2014_CIM_XX_Atti.pdf	Yes
35	His engine's voice: towards a vocal sketching tool for synthetic engine sounds	S. Baldan, S. Delle Monache and L. Comanducci	XX Colloquio di Informatica Musicale	Rome, Italy, October 20–22	AIMI	Venice, Italy	2014	134–139	http://smc.dei.unipd.it/cim_proceedings/2014_CIM_XX_Atti.pdf	Yes
36	A design exploration on the effectiveness of vocal imitations	S. Delle Monache, S. Baldan, D.A. Mauro and D. Rocchesso	40th International Computer Music Conference (ICMC) joint with the 11th Sound and Music Computing conference (SMC)	Athens, Greece, September 14–20	Michigan Publishing, University of Michigan Library	Ann Arbor, MI, USA	2014	1642–1648	http://smc.afim-asso.org/smc-icmc-2014/images/proceedings/PS1-B12-Adesignexploration.pdf	Yes
37	2001-2016: Oggetti sonanti in Europa	D. Rocchesso	41° Convegno Nazionale dell'Associazione Italiana di Acustica	Pisa, Italy, June 17–19	AIA		2014			No
38	Sound initiation and source types in human imitations of sounds	P. Helgason	Proceedings of FONETIK	Stockholm, Sweden, June 09–11			2014	83–88	http://www.diva-portal.org/smash/get/diva2:730213/FULLTEXT01.pdf	Yes
39	Sketch a Scratch	S. Delle Monache, D. Rocchesso, and S. Papetti	8th International Conference on Tangible, Embedded and Embodied Interaction (TEI)	Munich (Germany), February 16–19	ACM	New York, NY, USA	2014			No
36	Extraction of articulatory parameters from audio files used for sound sketching	M. Hellwagner	M.Sc. thesis	TU Berlin and KTH Stockholm			2016			Yes
37	Combination of gesture and vocalization in the imitation of sounds	H. Scurto	M.Sc. thesis	UPMC – IRCAM Paris			2015			Yes

2.1.3 List of dissemination and communication activities

TABLE A2: LIST OF DISSEMINATION ACTIVITIES

NO.	Type of activities	Main leader	Title	Date/Period	Place	Type of audience	Size of audience	Countries addressed
1	Conference	Davide Rocchesso	Sound Design Rendezvous	19 January 2017	IRCAM Paris	Scientific community Civil Society	60	International
2	Web interview	Patrick Susini	.Sound: A Material for Design	16 December 2016	https://www.ircam.fr	Civil Society	N/A	International
3	Seminar	Geoffroy Peeters	Recent researches at IRCAM related to the recognition of rhythm, vocal imitations and music structure	15 December 2016	Johannes Kepler University in Linz, Austria	Scientific community	N/A	International
4	Seminar	Geoffroy Peeters	Recent researches at IRCAM related to the recognition of rhythm, vocal imitations and music structure	18 November 2016	Pompeu Fabra University in Barcelona, Spain	Scientific community	N/A	International
5	Conference	Patrick Susini	Keynote Speech at Audiomostly	6 October 2016	Norrkp̈ing, Sweden	Scientific community	about 100	International
6	Exhibition	Stefano Baldan	Disegnare il suono con un microfono	30 September 2016	Venice, Italy	Civil society	about 100	National
7	Radio broadcast	Frédéric Bevilacqua	Talk about sound design at France Inter radio	18 September, 2016	Paris, France	Civil society	thousands	National
8	Workshop	Stefano Delle Monache	Vocal sketching at Acusmatiq Festival	29 July, 2016	Ancona, Italy	Practitioners	N/A	National
9	Lecture	Davide Rocchesso	Product Sound Design at Virtual Prototyping Summer School	13 July, 2016	Politecnico di Milano, Italy	Higher education	about 20	International
10	Seminar	Davide Rocchesso	Sounding Objects and Sonic Sketches	6 July, 2016	University of Palermo, Italy	Scientific community	about 20	National
11	Video	Sylvestre Miget	48h of sound design at Château La Coste	June, 2016	Paris, France	Civil society	352	International
12	Workshop and Exhibition	Patrick Boussard	48h of sound design at Château La Coste	28-30 April, 2016	Aix en Provence, France	Practitioners, Civil society	about 200	International
13	Presentation	Davide Rocchesso	Open SkAT-VG Lab for the World Voice Day	15 April, 2016	Iuav University of Venice, Italy	Scientific community, civil society	about 30	National
14	Conference and TV broadcast	Sten Ternström	The Voice as a Professional Tool, broadcast on Swedish national educational TV	16 April, 2016	KTH Stockholm, Sweden	Scientific community, civil society	thousands	National
15	Seminar	Davide Rocchesso	Imagining, sketching and prototyping sound	4 March, 2016	European Centre for Living Technology, Venice, Italy	Scientific community	about 30	International
16	Web article	Guillaume Lemaitre	Soon We Will be Able to Design Custom Sounds with Voice And Gesture	1 March 2016	gizmodo.com	Civil society	N/A	International
17	Web Article	Davide Rocchesso	FET Open project SkAT-VG decodes sounds behind human gestures, in FET Through the Keyhole	February, 2016	European Commission	Scientific community, civil society, policy makers	N/A	International
18	Web Video	Guillaume Lemaitre	What Sound Is This Gesture?, interview for Inside Science	February, 2016	insidescience.org	Scientific community, civil society	N/A	International
19	Conference	Patrick Susini	The Skat-VG project : a move to a new sound design tool	27 January, 2016	Ircam Paris, France	Scientific community, civil society	N/A	International
20	Press release	Guillaume Lemaitre	Combining gestures and vocalizations during sound imitation, lay-language version for online press room of ASA	November, 2015	Jacksonville, USA	Scientific community, civil society	N/A	International
21	Seminar	Davide Rocchesso	Imagining, sketching and prototyping sound, Sound and Music Computing Colloquium	26 November, 2015	Aalborg University Copenhagen, Denmark	Scientific community	about 20	International
22	Workshop	Stefano Delle Monache	Workshop on vocal sketching with students of the Master in Sound and Music Computing	23-27 November, 2015	Aalborg University Copenhagen, Denmark	Higher education	about 30	International
23	Exhibition	Davide Rocchesso	SkAT-VG at ICT 2015	October, 2015	Lisbon, Portugal	Scientific community, civil society, industry, policy makers	hundreds	International
24	Installation	Andrea Cera	S'i' fosse suono	October, 2015	Lisbon, Portugal	civil society	hundreds	International
25	Workshop	Stefano Delle Monache	Workshop on vocal sketching with students of Dept. Film, Theatre and Television	7-8 May, 2015	York University, UK	Higher education	about 12	International
26	Seminar	Davide Rocchesso	Sonic Interaction Design	11 December 2014	SUPSI Lugano, Switzerland	Higher education	about 20	International
27	Seminar	Patrick Susini	Sketching Audio Technologies using Vocalization and Gestures: Le projet SkAT-VG, Projet Européen FP7-ICT FET-Open: challenging current thinking	December 2014	Ircam Paris, France	Scientific community	about 30	International
28	Exhibition	Alan Del Piccolo	Sketch a Scratch, Opening of the Doctoral Year	14 November, 2014	Ca' Foscari University of Venice, Italy	Higher education	about a hundred	International

29	Press release	Davide Rocchesso	Uttering the sounds of future products, in English, Italian, and French	7 November 2014	Venice, Paris, Stockholm	Civil society	N/A	International
30	Workshop and Demos	Patrick Susini	Sketching Audio Technologies using Vocalizations and Gestures, Ircam open days	11 June, 2014	Ircam Paris, France	Scientific community, civil society	N/A	International
31	Seminar	Sten Ternström	SkAT-VG presentation for the World Voice Day	16 April, 2014	KTH Stockholm, Sweden	Scientific community, civil society	N/A	International
32	Performance	Davide Rocchesso	Textures from an Exhibition, for the World Voice Day	16 April, 2014	Iuav University of Venice, Italy	Scientific community, civil society	about 40	National
33	Seminar	Pétur Helgason	SkAT-VG presentation at the Marcus Wallenberg Laboratories for Sound and Vibration	4 April, 2014	KTH Stockholm, Sweden	Scientific community	N/A	International
34	Workshop	Stefano Delle Monache	Workshop on vocal sketching with students of the Conservatory	24 March, 2014	Padova, Italy	Higher education	about 20	National

2.2 Section B

The SkAT-VG project has produced several results that are worth considering for possible exploitation. Some of these results are functional to the incremental building of scientific knowledge, they are disseminated through scientific publications, and distributed through conventional scientific channels. In section 2.2.2 we only list four possible direction of exploitation, which may lead to several products, and we develop the exploitation plan for one of these Key Exploitable Results (KER). This exploitation plan is the result of the Exploitation Strategy Seminar (ESS) that was organized for SkAT-VG by the Common Exploitation Booster[2]. So far, no application has been made for patents, or trademarks, and therefore the table in section 2.2.1 is empty. However, the consortium member GENESIS already found a profitable way to exploit some findings of SkAT-VG in their LEA product and Active Sound Design process (see table 2).

2.2.1 Part B1

TABLE B1: L IST OF APPLICATIONS FOR PATENTS , TRADEMARKS , REGISTERED DESIGNS , ETC.

Type of IP Rights	Confidential (YES/NO)	Foreseen embargo date	Application reference	Subject or title of application	Applicant(s)(as on the application)

2.2.2 Part B2

The SkAT-VG project has produced a foreground that has potential for commercial exploitation as well as for general advancement of knowledge and social innovation. A number of Key Exploitable Results (KER) are listed in table B2.

TABLE B2: EXPLOITABLE FOREGROUND

n.	Type of exploitable foreground	Exploitable product(s) or measure(s)	Description of exploitable foreground	Confidential (YES/NO)	Foreseen embargo date	Sector(s) of application	Timetable, commercial or any other use	Patents or other IPR exploitation (licenses)	Owner and other beneficiary(s) involved
1	Commercial exploitation of R&D results	SEeD: Sonic Embodied Design	a Tool to set and play with sound models by voice and gesture	NO		R90	2018 for startup	to be defined	IRCAM and IUAV
2	Commercial exploitation of R&D results	SDT: The Sound Design Toolkit	Procedural audio synthesis tools for videogame and audio-visual product development	NO		R90	2018	GPL	IUAV
3	Commercial exploitation of R&D results	Imitation game	A board game based on vocal imitations of sounds	NO		C32.4	2018	to be defined	IUAV
4	General advancement of knowledge	MMDB	Multimodal Database of vocal and gestural imitations elicited by sounds	NO		M72.1	2017	Free use for research or education	IRCAM and KTH

The KERs are briefly described as follows:

SEeD tool for setting an ensemble of sound models and further playing with them by voice and gesture. It is a sketching tool for sound, designed for embodied interaction. It requires a limited amount of further research but an extensive development work, driven by requirements collected through experimentation and dialogue with stakeholders. Its exploitation plan is further illustrated below;

SDT is a set of sound synthesis and processing modules based on the physical description of processes and events. At its current stage of maturation it is suitable for being used in a large variety of projects and application areas, ranging from games to audiovisual

[2] http://exploitation.meta-group.com

products. It is released as a free software under the GNU Public License, which should favor its adoption among students, educators, developers and artists;

Imitation game is a board game for edutainment which proved to be effective to let workshop participants familiarise with vocal sketching. When fully developed as a product, it has potential to foster social activities based on sound communication;

MMDB is a multimodal database of recordings (audio, video, sensor data, computed features) of lay persons and professional actors producing imitations of referent sounds. A subset of the database is annotated by experienced phoneticians. It is a resource for research in psychoacoustics, phonetics, and sound sciences, and it is made freely available to scientists and educators.

Exploitation plan for SEeD

In fall 2016, the SkAT-VG project took advantage of the Common Exploitation Booster, a support service for facilitating exploitation of research results of ongoing research and innovation projects, under FP7 and H2020. An Exploitation Strategy Seminar (ESS) was run on November 2016 in Venice, Italy, with participants from IUAV, IRCAM, and GENESIS. The ESS is a brainstorming exercise to characterise the KERs, discuss key features, risks and obstacles for using them. The participants and the facilitator, Alessia Melasecche Germini, decided to focus on KER 1 of Table B2, the project outcome that is most likely to become a technological product. Overall, the experience of the seminar was very positive, as the project participants had little or no experience in innovation processes, and this opportunity was important to see how such processes can be tailored to a specific research result.

A detailed exploitation analysis has been conducted for SEeD (Sonic Embodied Design, KER 1 of Table B2), which moved from the construction of a Lean Business Model Canvas, as depicted in figure 16. A Characterization table is provided in figure 17, a Risk matrix with a Priority map is provided in figure 18. In the priority map, the top-left quadrant is for risks that are under control, while the top-right quadrant is for risks that require action.

SkAT-VG Consortium partners IUAV and IRCAM submitted the proposal ESD (Embodied Sound Design) to the first call of the FET Innovation Launchpad, for the development of a commercial product starting from the SEeD prototype. If the ESD proposal would have been funded, it would have been possible to tailor product development to markets capable of generating revenue and growth. The selection of funded Launchpad proposals was completed in the third week of January 2017, and unfortunately it didn't include ESD. However, SkAT-VG is still eligible to apply for the second call in 2017. The ESD proposal will be refined, taking advantage of knowledge gained from the Common Exploitation Booster, and submitted to this new call.

The main goal of ESD is to define an exploitation path, by conducting market and user studies informing the assembly and assessment of the SkAT-VG innovation. SkAT-VG identified sound designers from the video gaming, movie, and automotive industries as potential customers looking for intuitive tools to boost their creativity and productivity, but it also revealed other unexpected pathways for potential exploitation. For instance, workshops showed that embodying sounds by voice and gesture is an entertaining activity, suggesting interactive and educational games for the general public as another potential market for the same technology. The first line of work focuses on the main idea: embodied professional sound design applications for the video gaming, movie, and automotive industries. A second parallel line explores other opportunities, such as the entertainment and educational applications mentioned above, and others that may emerge along the way. A three-tiered implementation consists of primary market research, development and assessments of technological scenarios, and definitions of exploitation paths. The final outcome would be the identification of the most profitable markets, paving the way for the development of an innovative company. It would also turn some scientists of the SkAT-VG projects into European entrepreneurs.

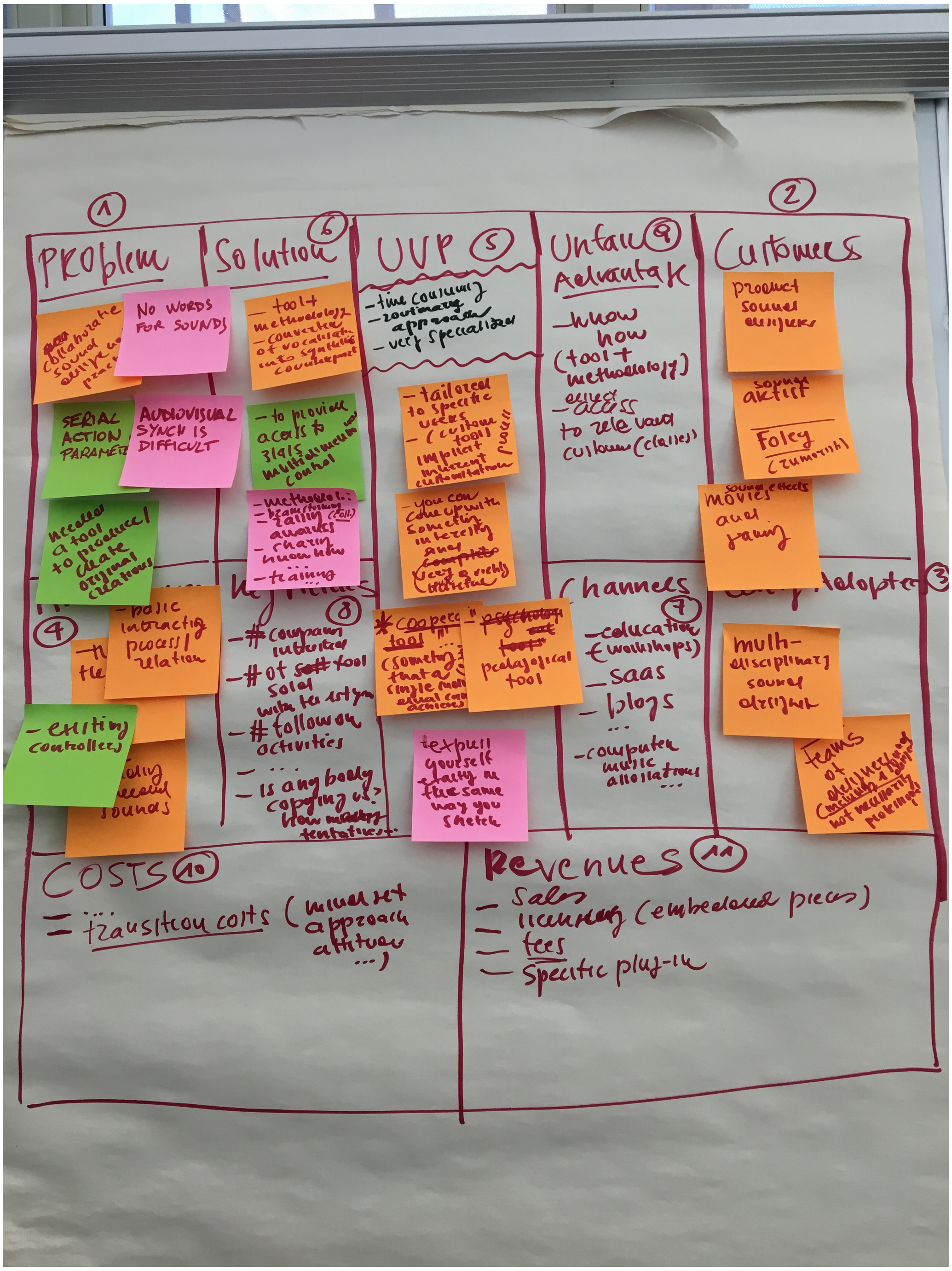

Figure 16: The Lean Canvas of KER 1 on a flipchart

KER 1	SEeD – Sonic Embodied Design
Description of the Result	Automatic system for the recognition, classification and real-time resynthesis of vocal imitations. A tool that allows the designer to first set physics-based or corpus-based sound models, and then to play interactively with them to create elaborate sounds
Problems you are addressing and how your customers solve them so far	Providing a yet missing tool for sketching in sound design. So far sketching in the aural domain has been a missing practice. Sound designers immediately jump to the final product without trying to explore different and innovative ideas first. The goal is to increase variety and speed of idea generation in sound creation.
Unique Selling Point	SEed affords sound sketching by imitation-driven synthesis: machine learning tools to recognize and classify vocal imitations, signal analysis algorithms to extract relevant information from vocal and gestural input, a library of physically-informed and corpus-based sound models for everyday sounds controllable in real time with voice and gesture.
Product/Service Market Size	Small B2C niche. Selling to individuals or studios.
Market Trends/Public Acceptance	Noticeable trend towards the adoption of procedural audio, especially in video games. Approach and prototype tools positively accepted by sound designers participating in SkAT-VG workshops.
Product/Service Positioning	Professional sound designers and artists.
Legal or normative or ethical requirements (need for authorisations, compliance to standards, norms, etc.)	Technology mostly released with open licenses. Proprietary parts belonging to the project partners (Ircam).
Competitors/Incumbents	Audiogaming offers procedural audio packages for the dynamic generation of several sound effects like rain, wind, and car engines. Krotos Dehumanizer provides audio effects to transform the human voice into a “creature” voice (robot, monster, etc.)
Early Adopters - First Customers	Prospects are sound design professionals who took part to SkAT-VG workshops
Cost of implementation - bringing product/service to the “market” (before Exploitation)	In the order of 100,000EUR for incubation of product and producing company
Time to market (from the end of the project)	18 – 36 months
Foreseen Product/Service Price	Under 500 EUR
Adequateness of Consortium Staff	Skills in software development, user-base creation, and market analysis are available or could be acquired at Iuav and Ircam
External Experts/Partners to be involved	
Status of IPR: Background (type and partner owner)	Ircam descriptors. Sound Design Toolkit (Iuav). GMM classifier (Ircam). Corpus-based synthesis (Ircam).
Status of IPR: Results/Foreground (type and partner owner)	Extension of Sound Design Toolkit, including alternative descriptors (Iuav). HMM classifier (Ircam). SeeD architecture and interface (Ircam and Iuav).
Status of IPR: use the results from the Exploitation Form	IPR will continue belonging to the partners while the idea is incubated through additional public funding.
Partner/s involved expectations	Partners expect sketching tools will increase popularity of sound creation to a large public of non-professionals.
Sources of financing foreseen after the end of the project	Possible 12-months incubation through FET Innovation Launchpad. The proposal is being evaluated.

Figure 17: Characterization of KER 1

	KER 1	Degree of importance of the risk related to the final achievement of this Key Exploitable Result. Please rate from 1 to 10 (1 low- 10 high)	Probability of risk happening Please rate from 1 to 10 (1 low - 10 high)	Risk Grade	Scope and type of potential intervention	Feasibility/Success of Intervention Please rate from 1 to 10 (1 low- 10 high)	Priority Level
1	**Partnership Risk Factors**	8	3	27		8	213
	Disagreement on ownership rules	8	6		Legal mediation of ownership disputes and patent review	9	
	Industrialization at risk: a partner declares bankruptcy.	8	2		initiate consortium meeting to explain usage issues and redefine roadmap of exploitation	6	
	Industrialization at risk: a partner leaves the market	8	2		maintain communication with exiting expert and organise hand-over and training	9	
2	**Technological Risk Factors**	8	5	40		7	280
	better technology emerges	8	5		Re-evaluation of technology and further optimisation to match/outperform new benchmark	7	
3	**Market Risk Factors**	5	5	23		7	146
	Exploitation disagreement	3	2		discuss exploitation proposals and ensure that parallel exploitation is beneficial and in the interest of marketing the technology	6	
	difficulty in market penetration/customer reception and acceptance of technology	7	7		additional market studies, customer surveys and assessment of product shortfalls	7	
4	**IPR/legal Risk Factors**	8	6	48		8	384
	competitors replicate technology	8	6		stricter control of technical know-how	8	
5	**Financial/management Risk Factors**	6	9	54		6	324
	weak exploitation of the material	6	9		revision of exploitation plan and market research and relaunch of product	6	
6	**Environmental/regulatory Risk Factors**	4	1	4		8	32
	not in compliance with regulations	4	1		assessment of legal/regulatory requirements and alteration of product to comply	8	

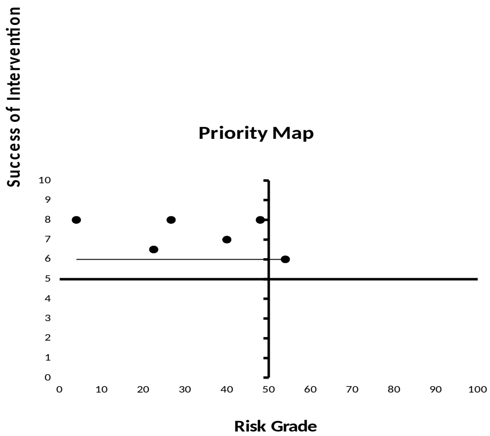

Figure 18: Risk matrix and Priority map for KER 1

3 Report on societal implications

A General Information *(completed automatically when **Grant Agreement number** is entered.*

Grant Agreement Number:	**FP7-ICT-2013-C FET-618067**
Title of Project:	**Sketching Audio Technologies using Vocalizations and Gestures**
Name and Title of Coordinator:	**Prof. Davide Rocchesso**

B Ethics

1. Did your project undergo an Ethics Review (and/or Screening)? • If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports? Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 *'Work Progress and Achievements'*	***√Yes 0No***
2. Please indicate whether your project involved any of the following issues (tick box) :	***YES***
RESEARCH ON HUMANS	
• Did the project involve children?	
• Did the project involve patients?	
• Did the project involve persons not able to give consent?	
• Did the project involve adult healthy volunteers?	√
• Did the project involve Human genetic material?	
• Did the project involve Human biological samples?	
• Did the project involve Human data collection?	
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	
• Did the project involve Human Foetal Tissue / Cells?	
• Did the project involve Human Embryonic Stem Cells (hESCs)?	
• Did the project on human Embryonic Stem Cells involve cells in culture?	
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	
PRIVACY	
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	
• Did the project involve tracking the location or observation of people?	
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	
• Were those animals transgenic small laboratory animals?	
• Were those animals transgenic farm animals?	
• Were those animals cloned farm animals?	
• Were those animals non-human primates?	
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	
• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	
DUAL USE	

• Research having direct military use	No
• Research having the potential for terrorist abuse	No

C Workforce Statistics

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).

Type of Position	Number of Women	Number of Men
Scientific Coordinator		1
Work package leaders		7
Experienced researchers (i.e. PhD holders)	5 4 (kth) + 1 (Genesis)	19 3 (iuav) + 5 (kth) + 9 (ircam) + 2 (Genesis)
PhD Students		2 1 (iuav) + 1 (kth)
Other	5 2 (iuav) + 1 (kth) + 1 (ircam) + 1 (genesis)	7 2 (iuav) + 2 (kth) + 2 (ircam) + 1 (genesis)

4. How many additional researchers (in companies and universities) were recruited specifically for this project?	**13**
Of which, indicate the number of men:	10

D Gender Aspects

5. Did you carry out specific Gender Equality Actions under the project?	✓	Yes No

6. Which of the following actions did you carry out and how effective were they?

	Not at all effective	Very effective
✓Design and implement an equal opportunity policy	✓	○
Set targets to achieve a gender balance in the workforce		○
Organise conferences and workshops on gender		○
Actions to improve work-life balance		○
Other:		

7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?

Yes- please specify

✓No

E Synergies with Science Education

8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?

✓Yes- please specify: M.Sc. Theses, World Voice Day, European Researchers' Night

No

9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?

✓Yes- please specify: Workshops with students; website

No

F Interdisciplinarity

10. Which disciplines (see list below) are involved in your project?

Main discipline[1]: 1.1

Associated discipline1: 1.2	○	Associated discipline1: 5.1

G Engaging with Civil society and policy makers

11a Did your project engage with societal actors beyond the research community? *(if 'No', go to Question 14)*		✓ Yes No

[1] Insert number from list below (Frascati Manual).

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?

No

Yes- in determining what research should be performed

Yes - in implementing the research

✓ Yes, in communicating /disseminating / using the results of the project

11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?	○ ○	✓ Yes No

12. Did you engage with government / public bodies or policy makers (including international organisations)

No

Yes- in framing the research agenda

Yes - in implementing the research agenda

✓ Yes, in communicating /disseminating / using the results of the project

13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?

Yes – as a **primary** objective (please indicate areas below- multiple answers possible)

✓ Yes – as a **secondary** objective (please indicate areas below - multiple answer possible)

No

13b If Yes, in which fields?

Agriculture **Audiovisual and Media** Budget Competition Consumers Culture Customs Development Economic and Monetary Affairs **Education, Training,** Youth Employment and Social Affairs		Energy Enlargement Enterprise Environment External Relations External Trade Fisheries and Maritime Affairs Food Safety Foreign and Security Policy Fraud Humanitarian aid		Human rights Information Society Institutional affairs Internal Market Justice, freedom and security Public Health Regional Policy **Research and Innovation** Space Taxation Transport	

13c If Yes, at which level?

Local / regional levels
✓ National level
✓ European level
International level

H Use and dissemination

14. How many Articles were published/accepted for publication in peer-reviewed journals?	**12**
To how many of these is open access[2] provided?	**9**
How many of these are published in open access journals?	**3**
How many of these are published in open repositories?	**9**
To how many of these is open access not provided?	**3**
Please check all applicable reasons for not providing open access:	
❑ publisher's licensing agreement would not permit publishing in a repository ❑ no suitable repository available ✓no suitable open access journal available ❑ no funds available to publish in an open access journal ✓lack of time and resources ❑ lack of information on open access ❑ other[3]: ……………	

15. How many new patent applications ('priority filings') have been made? *("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).*	**0**

16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	**0**
	Registered design	**0**
	Other	

17. How many spin-off companies were created / are planned as a direct result of the project?	**0**
Indicate the approximate number of additional jobs in these companies:	

18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:

Increase in employment, or	❑	In small & medium-sized enterprises
Safeguard employment, or	❑	In large companies
Decrease in employment,	❑	None of the above / not relevant to the project
✓Difficult to estimate / not possible to quantify		

[2] Open Access is defined as free of charge access for anyone via Internet.
[3] For instance: classification for security project.

19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (*FTE = one person working fulltime for a year*) jobs:	*Indicate figure:* *12*
Difficult to estimate / not possible to quantify	❑

I Media and Communication to the general public

20. As part of the project, were any of the beneficiaries professionals in communication or media relations?

Yes ✓No

21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?

✓Yes No

22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?

✓Press Release	✓	Coverage in specialist press
Media briefing	✓	Coverage in general (non-specialist) press
✓TV coverage / report	✓	Coverage in national press
✓Radio coverage / report	✓	Coverage in international press
✓Brochures /posters / flyers	✓	Website for the general public / internet
✓DVD /Film /Multimedia	✓	Event targeting general public (festival, conference, exhibition, science café)

23 In which languages are the information products for the general public produced?

Language of the coordinator	✓	English
Other language(s)	✓	Italian, Swedish, French

***Question F-10*:** Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES
1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]
1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)
1.3 Chemical sciences (chemistry, other allied subjects)
1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)
1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY

2.1 Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)

2.2 Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]

2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; the science and technology of food production; specialised technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES

3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)

3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)

3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES

4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)

4.2 Veterinary medicine

5. SOCIAL SCIENCES

5.1 Psychology

5.2 Economics

5.3 Educational sciences (education and training and other allied subjects)

5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary , methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]