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

SkAT-VG: Sketching Audio Technologies using Vocalizations and Gestures

D1.1.2
Periodic Report

First Author	Davide Rocchesso
Responsible Partner	IUAV
Status-Version:	Draft-1.0
Date:	January 5, 2017
EC Distribution:	Consortium
Project Number:	618067
Project Title:	Sketching Audio Technologies using Vocalizations and Gestures

Title of Deliverable:	Periodic Report
Date of delivery to the EC:	09/01/2017

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 Periodic Report addresses the technical aspects of the project in its third and final year. Previous periods are covered by D1.1.1 and D1.1.1bis.
Keyword List:	periodic report

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 D1.1.2

Version	**Date**	**Description**	**Author**
First draft	03/11/2016	Import from template	ROC
First draft	25/11/2016	Updating WP7	CDe
First draft	26/11/2016	Updating WP4	GL
Complete draft	25/12/2016	Major pass	ROC
Final	05/01/2017	Final touch	ROC

Project Periodic 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

Periodic report: P2 (third year)

Period covered: from 01 January 2016 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

Declaration by the Scientific Representative of the Project Coordinator

I, as scientific representative of the coordinator of this project and in line with the obligations as stated in Article II.2.3 of the Grant Agreement declare that:

The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;

The project (tick as appropriate):

- ☑ has fully achieved its objectives and technical goals for the period;
- ☐ has achieved most of its objectives and technical goals for the period with relatively minor deviations.
- ☐ has failed to achieve critical objectives and or is not at all on schedule.

The public website, if applicable

- ☑ is up to date
- ☐ is not up to date

To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project (Section 4) and if applicable with the certificate on financial statement.

All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes have been reported under Section 2.4 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of scientific representative of the Coordinator: Davide Rocchesso

Date:

Signature of scientific representative of the Coordinator: .....

Table of Contents

Index of Figures

List of Acronyms and Abbreviations

DoW Description of Work

EC European Commission

PM Person Month

WP Work Package

GA Grant Agreement

CA Consortium Agreement

M Milestone

Mo Month

Q Quarter

1 Publishable summary

Sketching Audio Technologies using Vocalizations and Gestures
www.skatvg.eu

Università Iuav di Venezia

Pleasant, yet functional. The fusion of these two adjectives describes most of the work of designers, in any domain. In the aural 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 idea of SkAT-VG is to exploit the most natural of sound design tools: human voice and gestures. Humans have surprising capabilities in communicating sound, especially in interactive contexts, but a thorough understanding of how this happens and how these capabilities could be exploited required an ambitious research plan. In the project, the Iuav University of Venice has been developing design methods and tools based on vocal and gestural sketching. The French company Genesis provided an industrial framework and application contexts. The possibilities and limitations of the human voice as a sound sketching apparatus have been charted by partner KTH in Stockholm, where thousands of utterances have been collected, annotated, and classified in relation to the physical phenomena they are supposed to mimic. This work has been done together with the Ircam institute in Paris, where three research teams have been involved in studying the features (Sound Analysis-Synthesis), understanding the human response (Perception and Sound Design), and exploiting the non-acoustic component (Sound Music Movement Interaction) of imitative or evocative vocalizations.

The four institutional partners of SkAT-VG have been interacting with professional and academic stakeholders in order to define the scope of sound design in future interactive contexts. In the near future, designers might sketch novel responsive sounds, for a car or for a coffee machine, by using the whole expressive potential of their voice and body.

1.1 Context and Objectives

Sketches are materials for the development and communication of ideas in the early stages of any design process. Sketching is commonly done with paper and pencil in visual design, but it is far from being straightforward when it comes to drafting the sonic behavior of objects. SkAT-VG aimed at understanding and exploiting vocalizations and gestures, which are the most natural analogues to hand and pencil to communicate sound in action.

SkAT-VG aimed at achieving three main objectives, which will lead to improved understanding, classification, and exploitation of human vocalizations and gestures:

1. **Understand.** To extend existing knowledge in perception and production of vocal imitations and expressive gestures;

2. **Classify.** To develop automatic classifiers of vocal and gestural imitations, based on what is imitated, by integrating signal analysis with the physio-mechanics of vocal production;

3. **Design.** 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.

These objectives define the three SkAT-VG Milestones:

M1 Accumulation of a large enough database of recorded, sorted, and labeled imitations (Mo12-15);

M2 Automatic classifiers of vocal and gestural imitations into categories of imitated sounds (Mo22-25);

M3 Integrated sketching tools (Mo36-37).

1.2 Description of Work

Most of the work in the first year of the project was aimed at achieving M1. In order to gain a better understanding of non-verbal communication through voice and gesture, the project partners KTH and IRCAM coordinated a campaign of recordings, measurements, and experiments with human participants. At KTH attention was more focused on extracting the primitives of vocal production, as they emerge from non linguistic tasks. A special representation for articulators was devised and used to annotate the recordings. Conversely, IRCAM moved from everyday sounds to see how people organize them into perceptual spaces and use these internal representations to communicate by means of vocal and gestural imitations. Expert and naïve performers were asked to produce vocalizations that vary along elementary auditory features, and the results showed that the fidelity of the imitations is moderate for isolated acoustic features yet the identification of complex sound events is effectively conveyed by vocal imitations.

Milestone M2 has been achieved in the second year. In particular, IRCAM devised new features and developed original signal-based classifiers for vocal and gestural imitations. KTH

showed that articulatory features can be automatically extracted from audio by application of auditory receptive fields. New scientific knowledge has been emerging out of the convergence of three scientific perspectives: Production (Articulatory descriptors); Perception (Sound categories); Signals: (Signal primitives).

Progress towards M3 has been centered around interactions with sound design professionals and workshops on vocal sketching (GENESIS, IUAV). These activities have been exploiting the new scientific knowledge through design exercises and prototype systems for vocal and gestural sketching. Two major software frameworks have been produced and distributed, for physics-based sound synthesis and for the integration of audio modules in a sound design workflow. In the third year, IUAV and IRCAM have been collaborating at the realization of a unified vocal sketching tool, that makes physics-based and corpus-based sound synthesis readily available for rapid production of complex sound sketches. Such tool has been evaluated in sessions with professionals.

Aside from progress towards the milestones, scientific research has made progress toward a better understanding of vocal and gesture imitations. Studies have confirmed that listeners can recognize imitations very easily, and have analyzed how vocal and gestural imitations encapsulate different pieces of information important for sound recognition.

1.3 Main Results

The main achievements during the whole 36 months of SkAT-VG have been:

An exploratory database of audio recordings of various imitations, indexed and annotated by action primitives and attributes of origin [Hel14]. Recordings include diverse collected media, bespoke pilot recordings of three subjects, and an indexed set of sounds from the literature [New04]. Deliverable D2.2.1;

A database of recordings of skilled imitators, acquired with a structured and strict protocol, presented in detail in Deliverables D2.2.2 and D3.3.1;

Scientific evidence that vocal imitations communicate sounds more effectively than verbalizations [LR14];

A corpus of 52 referent sounds that are unambiguously identified (through an identification experiment) as belonging to three families and 26 categories: Included in Deliverable D4.4.1 (Mo21);

A setup for audio-visual recording of imitations of the referent sounds and pilot studies aimed at defining the recording protocol: Included in Deliverable D4.4.1 (Mo21);

A database of about 8000 recordings of vocal and gestural imitations, including audio recordings, high speed video recordings, depth camera and body tracking, and data from inertial measurement units attached on participants' wrists: Included in Deliverable D4.4.1 (Mo21). This database has been cleaned up and analyzed with acoustical

descriptors (Deliverable D5.5.1) and merged with the two KTH databases (Deliverables 3.3.1 and 3.3.2) in order to be accessible across the Internet[1];

A novel method for visualizing associations between large numbers of disparate tokens [PBO+14] has been refined and applied to the SkAT-VG database, in a collaboration with the KTH department for Computational Science and Technology. In order to enable the statistical analysis and visualization of articulations vs. referent sounds, the complete annotations of the skilled imitators (Swedish) and four lay imitators (French) were translated into a time-normalized format. This now allows us to see the degree of correspondence between, for example, types of voice sounds and types of referent sounds. Some examples are given in Deliverable D3.3.3.

A study was performed of imitator preferences for "tame" (linguistic) versus "wild" (acoustic) imitations. Conditions were: the recipient is 1) a human who speaks the imitator's language, 2) a human who does not understand the imitator's language, 3) a computer with audio imput only. The main findings were a) that the imitators produced more tame imitations when the recipient was a human who spoke their own language than for the other two conditions, and b) the morphology of referent sounds strongly affects the imitation preferences on the tame-wild continuum. (Deliverable D3.3.3). A manuscript is in preparation.

An analysis of vocal techniques and strategies used by lay and expert imitators to imitate basic acoustic features. It shows that imitators either reproduce the features faithfully or transpose them to the human voice range. This work was described in detail in D4.4.2 (Mo26) and has been presented at a conference [LJH+15] and published as a journal article [LJM+16];

A study that demonstrates that vocal imitations of most everyday sounds can be identified, even when listeners are not aware of the referent sound that is imitated (i.e. vocal imitations elicit semantic representations of what they imitate). Vocal imitations were compared to computational auditory sketches created by sparsifying time-frequency representations. The study was initially described in detail in D4.4.2 (Mo26) and has been presented at a conference [LHV+16] and a journal article [LHV+17];

A study of how imitators combine vocalizations and gestures to imitate sounds. The results showed that gestural imitations are inherently different from vocal imitations and highlighted iconic gestures used to communicate several acoustic properties of the sounds. The study was initially described in detail in D4.4.2 (Mo26) and has been presented at two conferences [SLF+15, SLF+16] and a journal article is in preparation [LSFB].

The application of a new method to identify which features of a vocal imitation are important for identification: a variation of the reverse correlation method labelled "auditory bubbles". A proof of concept of this method has been presented to a conference [ISL16];

[1] https://www.ircam.fr/project/blog/multimodal-database-of-vocal-and-gestural-imitations-elicited-by-sounds/

A large comparison of various strategies to perform automatic classification into imitation categories from the analysis of audio temporal series (vocal imitation) including continous/discrete hidden Markov model, morphological descriptors and dynamic time warping [MP15]: Included in Deliverable D5.5.1 (Mo23);

An informed classifier into imitation categories based on the automatic transcription of vocal imitations into vocal primitives: Included in Deliverable D5.5.2 (Mo28);

A late fusion classifier into imitation categories which combines the transcription into vocal primitives and the morphological descriptors: Included in Deliverable D5.5.2 (Mo28);

A novel approach to automatically estimate audio primitives given a dataset using Shift-Invariant Probabilistic Latent Component Analysis (SI-PLCA) applied to a constant-Q-transform: Included in Deliverable D5.5.2 (Mo28);

A novel approach that uses the automatically estimated audio primitives to represent imitation categories by a set of hidden Markov models: Included in Deliverable D5.5.2 (Mo28);

A novel movement representation, based on wavelet analysis and particle filtering, partially implemented under the MuBu framework: Included in Deliverable D5.5.1 (Mo23);

A set of analysis tools for voice analysis using Receptive Fields [LF15a][LF15b]: Included in Deliverable D5.5.1 (Mo23);

A software framework named SkAT Studio, for the integration of modules within custom audio processing workflows [BDRH16, HML+16]: Included in Deliverable D6.6.1 (Mo24);

Physics-based sound model of vehicle motor sounds [BLDB15]: Included in Deliverable D6.6.1 (Mo24);

A new software architecture and public release of the Sound Design Toolkit (SDT) framework [BDR17], integrated with all the sound models necessary to synthesize the timbral families emerged from the 26 perceptually discriminable categories of sounds: Included in Deliverable D6.6.1 (Mo24, revised Mo36);

A set of interviews with sound designers from France, Italy, UK, and Japan: Included in Deliverable D7.7.2 (Mo36). A journal article is in preparation [DBM+17];

Five workshops on vocal sketching, and a set of basic design exercises and activities on vocal and gestural sketching, arranged in a structured and modular workshop format [DBMR14, DRBM15, DR16a, DR16b, ERDS16]: Included in Deliverable D7.7.2 (Mo36);

The "48 hours of sound design at Château Lacoste" event, a SkAT-VG workshop in which five professional sound designers exploited the SkAT-VG tools (SkAT Studio, miMic and MIMES) to sonify five pieces of art. Included in Deliverable D7.7.2 (Mo36).

An industrial case study held with PSA (French car manufacturer), in which industrial sound designers could experience the SkAT-VG methodology, and test the tools: Included in Deliverable D7.7.2 (Mo36).

Sketch-a-Scratch, a tool for multisensory texture exploration at the tip of the pen [DRP14, RMP16, DDR+15]: Included in Deliverable D7.7.1 (Mo36);

miMic, a demonstrator based on the metaphor of the microphone as pencil. It is a system architecture that, through a microphone augmented with inertial sensors, can empower the user with a wide sound palette that can be directly controlled by voice and gesture [RDA16, HML+16]: Included in Deliverable D7.7.1 (Mo36);

MIMES, a family of 3D printed, interactive objects to support the vocal and gestural sketching of expressive sounds interactively [HML+16]: Included in Deliverable D7.7.1 (Mo36);

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: Included in Deliverable D7.7.1 (Mo36). A conference paper has been submitted [RBB+17], to be extended to a journal paper;

"S'i' Fosse Suono" installation, an interactive video mosaic of audio self-portraits [CMR16], which was developed in cooperation with the sound designer Andrea Cera, and evaluated with lay listeners. The installation exploits the concepts and tools being developed in the project, to communicate the long-term vision of SkAT-VG. Included in Deliverable D7.7.1 (Mo36).

Enlargement of the SkAT-VG community by establishing contacts and networking with interested external stakeholders, designers, scientists, and professionals (section 5 of this report);

Dissemination through public initiatives (World Voice Day 2014, ICT 2015, World Voice Day 2016, Researchers' Night 2016), seminars, press releases, and articles (Interactions, The New Soundtrack, ASA Press Room);

1.4 Expected Final Results and their Potential Impact and Use

The long-term vision of the SkAT-VG project is to introduce non-verbal vocalizations and expressive manual gestures at every stage of the sound design process, from early sketches to the final evaluation of the sound quality of products, wherever the sonic behavior of objects is relevant for their use and aesthetics.

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. 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. Although the SkAT-VG methods and tools have been introduced to a network of professionals and experimented in workshops and design sessions, it is too early to see a significant impact in the sound design community.

The outcomes of the SkAT-VG project are also aimed at fostering collaborative work in teams of professionals reflecting on sound problematics. Today, most of the sound design

activities involve a process of personal and internal creation. By bringing vocal and gestural sketching tools that are comprehensible, SkAT-VG aims at making a whole team working together, creating sounds together, at the beginning of a sonic project. Interactions between different actors are valuable in creative phases, and 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.

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 game analogue to Pictionary has been fully developed and tested in several workshops) and aids for language teaching (the recent experiment carried out at Språkstudion of Stockholm University and documented in D3.3.3 highlighted the possibility of using vocal imitations to help teaching foreign languages).

2 Core of the report for the period

The SkAT-VG project passed two intermediate technical reviews. The Addendum to the First Periodic Report, named D1.1.1bis and produced in October 2015, addressed the recommendations and comments raised by the reviewers after the review held in Venice on January 30, 2015. In this Second Periodic Report, we include replies to the critical comments that were given in the second Technical Review Report, held in Lisbon on October 23, 2015.

2.1 Response to Reviewers

These notes summarize the actions taken to address the observations made by the reviewers in their Technical Review Report dated February 3, 2016.

Issue: 1. – We continue to feel some concern with regards to the project focus, and we would like to encourage the consortium to continue concentrating on extracting the core ideas of the overall project, and make sure that the project will deliver convincing and high quality outcomes related to these core ideas.

Reply: The design and development of the SEeD tool (see section 2.3.7), an evolution and merge of the previously introduced miMic and MIMES tools, made some of the project research directions converge, and strengthened the collaboration between IRCAM and IUAV. The conduction and analysis of design sessions has been particularly important in the third year, and has been achieved through cross-partner collaboration. An important core idea that has been put forward by SkAT-VG through its scientific publications is that vocal sketching can be extremely useful as a probe to investigate auditory perception in general.

Issue: 2. – We would like to encourage further cross consortium collaborations. It is desirable for the consortium to demonstrate that prototypes and other outcomes are designed, implemented and delivered in such a way that each result informs the following, and that the core concepts are clearly seen to be commonly held and worked upon. In this context, an informal common methodological framework might become helpful. Other possible ways to do this will be the creation of cross-consortium co-authored publications, shared prototypes and co-hosted events and tests.

Reply: Continuing on the response to issue 1. IRCAM and KTH shared their databases of recordings and articulatory annotation system, and classifiers were trained based on this common ground. IUAV and GENESIS co-organised an industrial case study at PSA in Vélizy, France, the 7th and 8th November 2016, to make PSA team experiment SkAT-VG tools and methods (5 people: 1 sound design project manager, 2 audio-digital engineers, and 2 sound designers). IUAV and IRCAM collaborated to evaluate *S'i' fosse suono* and the SEeD tool with lay listeners and expert designers. Cross-consortium co-authored publications are [RBB+17, BDRH16, HML+16, DBM+17].

Issue: 3. – The third recommendation addresses exploitation. We would like to see the SKAT-VG tool evolve into something that is not only consistently used (and tested) in the work process of the project itself, but could also constitute a major part of the exploitation potential for the project.

Reply: In the last semester of the SkAT-VG project, significant efforts have been directed towards exploitation. In particular, the consortium applied for an Exploitation Strategy Seminar, offered by the Common Exploitation Booster, which was held in Venice on November 15, 2016, with participants from IUAV, IRCAM, and GENESIS. Based on the design and development of SEeD, a proposal named Embodied Sound Design (ESD) was submitted by IUAV and IRCAM for the FET Innovation Launchpad. It aims at the development of commercial products enabling embodied control of sound creation. It is currently under evaluation.

Issue: 4. – Progress towards project objectives: The project has achieved all the objectives from the DoW and has additionally delivered positive responses to the recommendations and concerns from the first review meeting. We remain somewhat concerned over the lack of a strong dissemination and exploitation plan.

Reply: The Final Report, which is being submitted simultaneously with the present Progress Report, contains the section "Use and dissemination of foreground" which contains both a dissemination strategy (sec. 2.1) and a plan for exploitation (sec. 2.2).

Issue: 5. – Collaboration and communication: Compared with the first review meeting, many other meetings have been organized and communication still proceed well. However, a stronger collaboration between the different partners is encouraged in writing joint scientific papers related to the project.

Reply: Cross-partner activities such as the PSA industrial case study or the SEeD evaluation sessions provided material that is being analyzed and digested into publishable form. Some cross-consortium publications already appeared (see reply to Issue 2) and others will follow.

2.2 Project Objectives for the Period

The SkAT-VG project has the following three main objectives:

DoW:

1. To extend existing knowledge in perception and production of vocal imitations and expressive gestures;
2. To develop automatic classifiers of vocal and gestural imitations, based on what is imitated, by integrating signal analysis with the physio-mechanics of vocal production;
3. 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.

As by the DoW, the objectives for the **third year** are summarized as

Understanding the role of linguistic aliases and biases in the communication of non-vocal sounds;

Fusing and integrating informed and blind classifiers of vocal and gestural imitations;

Developing design tools that enable sound creation and sound manipulation with the voice and the hands;

Evaluating how the design tools improve and facilitate sound creation and manipulation;

Exploring application areas for SkAT-VG tools.

Together, the above objectives lead to M3, scheduled for the end of the third year "Integrated sketching tools".

2.2.1 Structure of the work packages

The work plan is divided into work packages (WPs), which follow the logical phases of the implementation of the project, and include consortium management (in WP1) and assessment of progress and results (distributed among WPs and particularly relevant in WP7). Dissemination and exploitation are not described as a separate work package but rather distributed into the research work packages. Work packages are further decomposed into tasks, whose verifiable outcomes are called deliverables. The structure of WPs and Tasks, as well as their foreseen unfolding in time, is represented in Figure 1.

The development of the project over three years foresees three Milestones.

DoW:

Milestone M1 (Accumulation of a large enough database of recorded, sorted, and labeled imitations) represents the point where prior studies are integrated with new experimental results to provide enough accumulated data to effectively start the beginning of WP5 (automatic imitation recognition).

Milestone M2 (First implementation of automatic classifiers of vocal and gestural imitations) is the point where the first tools for segmentation and classification of imitations, informed by the knowledge in vocal and gestural production, will be available. This will boost the tasks in WP6 and WP7.

Milestone M3 (Integrated sketching tools) represents the final achievement of the project, where a tool to convert vocal and gestural sketches into instances of sound models will be available and evaluated. Also, the utility of vocal sketching in real-world design contexts will be manifest at this point.

The first period was mainly devoted to the initial fundamental scientific investigations on the perception and production of imitations, as well as to the analysis of requirements and definition of scenarios for applications of vocal and gestural sketching. These studies led to M1 at the end of Year 1. Milestone M2 was achieved by the end of Year 2 thanks to the studies on machine learning, and to the extensive analysis of vocal and gestural imitations. Major decisions on how to turn technologies into tools and applications (milestone M3) were taken already in the second year, and were adjusted and finalized in the third year.

The temporal organization of activities over the three years is represented in the PERT chart of Figure 2, where activities are assigned to branches, and the nodes represent significant stages of the project, including Milestones. This representation provides a concise view of the tasks and work packages as they were envisioned to overlap and unroll in time. Each elliptical node summarizes a period of the project (range of months) and it may contain a Milestone. The arcs are labeled with WPs or task numbers.

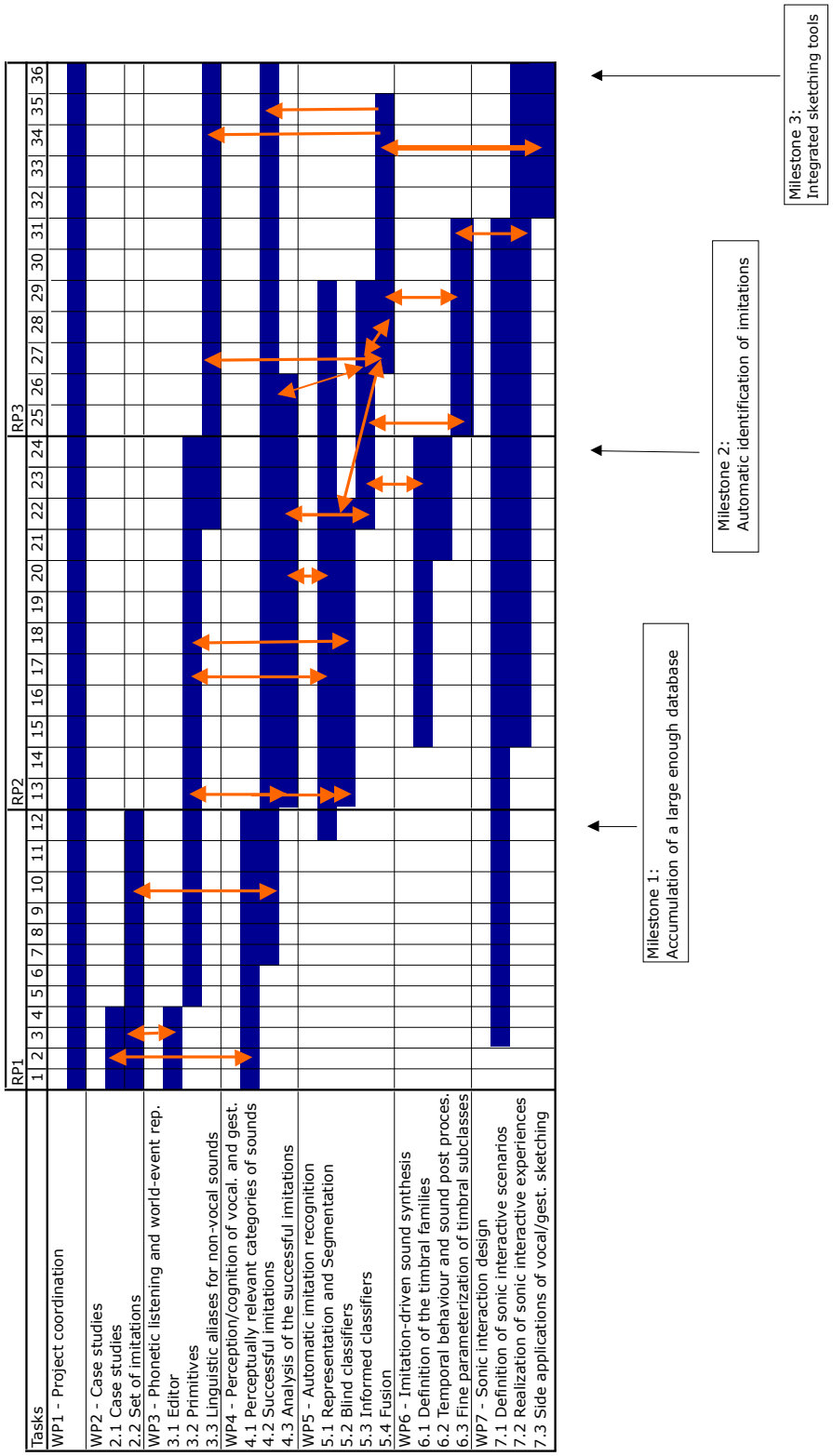


Figure 1: Timing of the different work packages as foreseen in the DoW. Orange arrows indicate interaction between WPs.

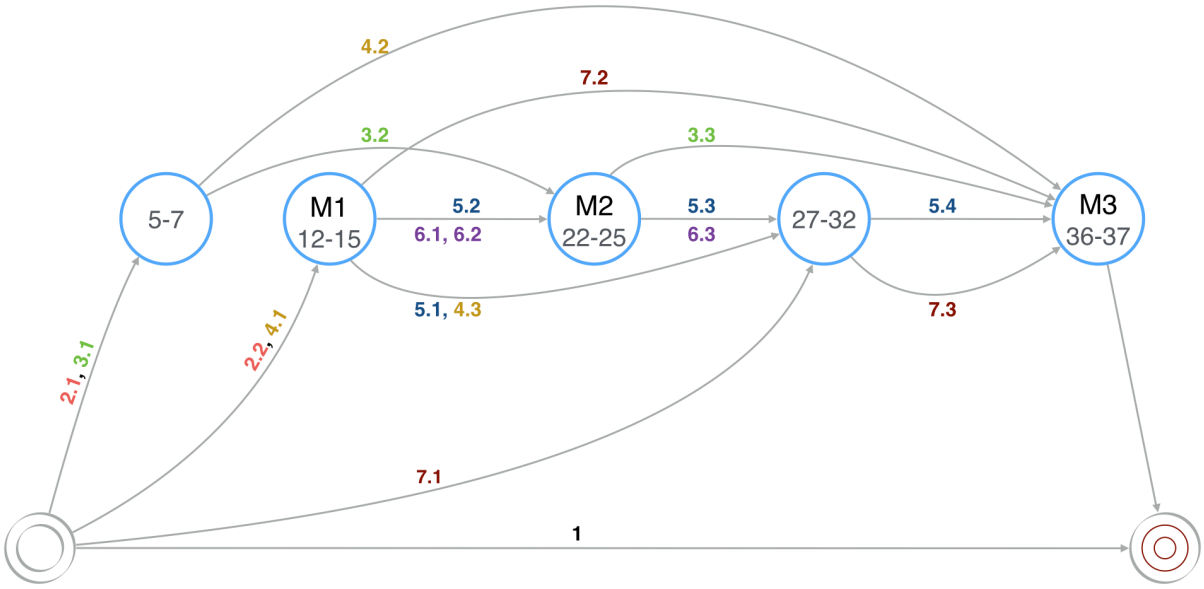


Figure 2: Interaction between the work packages.

2.2.2 Context of the Reporting Period within the Work Plan

Most of the activities in the Reporting Period are finalized at achieving Milestone M3:

WP3 (Phonetic listening and world-event representations): Task 3.2 (primitives) and Task 3.3 (linguistic aliases). Within the scope of Task 3.2, WP3 has provided WP5 with a modified version of the articulatory data set, conflating specific articulatory parameters and converting the data to a time series format. In collaboration with the KTH department for Computational Science and Technology, WP3 has worked to refine methods of data visualization developed by their team and to apply these methods to analyse the quite complex articulatory data set. For Task 3.3, WP3 has conducted a study in the use of "tame" vs. "wild" imitations in different contexts, focusing on the recipient. The main finding is that imitators are much less prone to use tame imitations when the recipient does not share a language with the imitator (human or computer), while tame imitations are used far more frequently when the recipient is a human who shares the imitator's language. The study also shows that the propensity for tame imitations is, to a large extent, a function of the acoustic morphology of the referent sound being imitated.

WP4 (Perception and cognition of vocalizations and expressive gestures): Tasks 4.2 (successful imitations) and 4.3 (analysis of successful imitations). During the reporting period, WP4 has consolidated the studies and analyses reported in D4.4.2: analysis of vocal strategies used by imitators, the demonstration that vocal imitations allow recognition of what is imitated, analysis of the combination of vocal and gestural imitations. The results of this consolidation consist of the presentation of the SkAT-VG research at several conferences and the publication of journal articles. A new study has also been started to highlight the acoustic features of imitations subserving identification.

WP5 (Automatic imitation recognition): Task 5.1 (representation and segmentation), Task 5.3 (informed classifiers) and Task 5.4 (fusion). During the third year, an informed classifier (based on the transcription of vocal imitations into vocal primitives) and a late-fusion classifier (based on the transcription of vocal imitations into vocal primitives and using morphological descriptors) have been developed. A real-time classification system has been developed for integration into prototype applications. A novel approach to automatically derive audio primitives from a given dataset (using Shift-Invariant Probabilistic Latent Component Analysis) has been proposed and used inside a HMM framework to recognize imitation categories.

WP6 (Imitation-driven sound synthesis): Task 6.3 (fine parameterization of timbral subclasses). Research in the definition of timbral families (Task 6.1) and temporal control of the synthesis algorithms (Task 6.2) has been put into practical use through the creation of two prototype environments for sonic sketching: miMic and Mimes. In the first one, vocal descriptors and physically informed sound synthesis models available in the SDT have been fine tuned to allow sonic sketching by imitation-driven sound synthesis. The second one mostly explores the use of gestures, enabling timbral manipulation of sounds (including vocal imitations) through physical manipulation of tangible objects and corpus-based concatenative sound synthesis.

WP7 (Sonic interaction design): Task 7.1 (definition of scenarios and applications); Task 7.2 (Realization of interactive experiences using the SkAT-VG Tools); Task 7.3 (Side applications of vocal and gestural sketching). During the third year, WP7 design research focused on consolidating the set of basic exercises and training around the elicitation, articulation and communication of complex sonic concepts via vocal imitations, and distilling them in a structured workshop format on vocal and gestural sketching. The resulting workshop format was instrumental to set a controlled playground through which informing the prototyping, exploration, and evaluation of the SkAT-VG tools: The two demonstrators miMic and Mimes were merged in a shared architecture "SEeD (Sonic Embodied Design)", based on the general framework of the third demonstrator SkAT Studio. In this respect, two main workshops involving expert sound designers focused on assessing the vocal and gestural sketching methodology, and the effectiveness of the SkAT-VG prototypes as mediation tools to support creativity in sound design tasks. Qualitative observations and quantitative evaluation based on protocol and linkographic analyses highlighted the value of cooperation in creative work, as one of the main SkAT-VG contributions to the discipline of sound design. In parallel, WP7 explored possible applications of vocal and gestural sketching, based on the feedbacks collected from the workshops and the interviews with sound practitioners. A new LEA module to enable spectral processing through gestural drawing on tablet has been wireframed and a first prototype was delivered.

Impact-generating activities

DoW: The SkAT-VG project has the potential to have an impact in science (understanding how vocal and gestural imitations are produced and perceived), technology (design tools, auditory displays, and sonic interactive artefacts), and society (designing a better soundscape for human beings and technological artefacts).

In its three years of activity, the SkAT-VG project has been producing new scientific knowledge, which actually represent the state of the art on the understanding how vocal and gestural imitations are produced and perceived: The classification on vocal production and articulation mechanisms is a foundation work, which goes well beyond the alphabet system of phonetic notation (IPA); The emerging ontology of perceptually relevant sound categories, and the identification of vocal and gestural primitives for the communication of sonic concepts, are at the edge of current research on perception and representation of sound. The collection of basic sound models, arranged in timbral families and made available on different platforms, represents an important step towards the design of physics-based sound synthesis tools, beyond the lab simulation of everyday sound phenomena. Starting with the realization of an artistic installation ("S'i fosse suono", by Andrea Cera), and going through the development of an embodied sound design tool (SEeD), both the general public and a niche of professionals could appreciate the SkAT-VG methodology and tools, and have a glance at future sound design practices.

The two concrete and external uses of SkAT-VG tools and methods ("The 48h of sound design" artistic workshop, and the "SkAT-VG / PSA case study") were crucial to establish an initial community of professional sound designers who are expected to become early adopters and evangelists of the SkAT-VG approach to sound design. Although sometimes part of the same team, sound designers were previously not used to create sounds together. From these two events and other workshops, the benefits of collaboration in sound design became evident to several stakeholders. The sessions of collaborative sound creation brings a real turning point in the sound design process.

SkAT-VG has been expanding the network of external stakeholders and expert sound designers, in order to tailor the design of sketching tools. In addition, educational research in workshop activities is involving students and professionals towards the development and systematization of sonic sketching exercises and practices. Finally, SkAT-VG dissemination strategy through press releases, publications in scientific magazines and journals, and public events has been strengthening the attention toward the SkAT-VG project, both in scientific communities and in the general public.

2.2.3 Project Objectives for the third year

The objectives for the third year of SkAT-VG are given in Table 1. It includes the declared Milestone and Deliverables as well as other objectives indicated in the DoW.

Milestones	Deliverables	Other
M3 - Integrated sketching tools	D3.3.3 - Report on how non-vocal world events are represented phonetically in SkAT-VG	Publications
	D4.4.2 - An analysis of how vocal and gesture primitives are sequenced	Publications
	D5.5.2 - Informed classifiers of imitations	Tools
	D5.5.3 - Integrated system that predicts the category of imitated sound sources	Tools
	D6.6.2 - Front-end applications for interactive sound prototyping	Tools / Publications
	D7.7.1 - Interactive prototypes realized with the SkAT-VG tool	Demonstrations
	D7.7.2 - Applications of vocal sketching	Workshops / Publications

Table 1: Objectives of SkAT-VG for the third year.

2.3 Work Progress and Achievements during the Period

The realization of SkAT-VG tasks and activities is represented in Figure 3, which has been automatically generated by the project-management tool.

Table 2 concisely shows the progress of WPs and Tasks in Years 1, 2, and 3, with the corresponding deviations from the planned progress. The deviations in each task are explained in the rest of this section. Table 3 assigns the Year 3 deviations to individual partners of the Consortium.

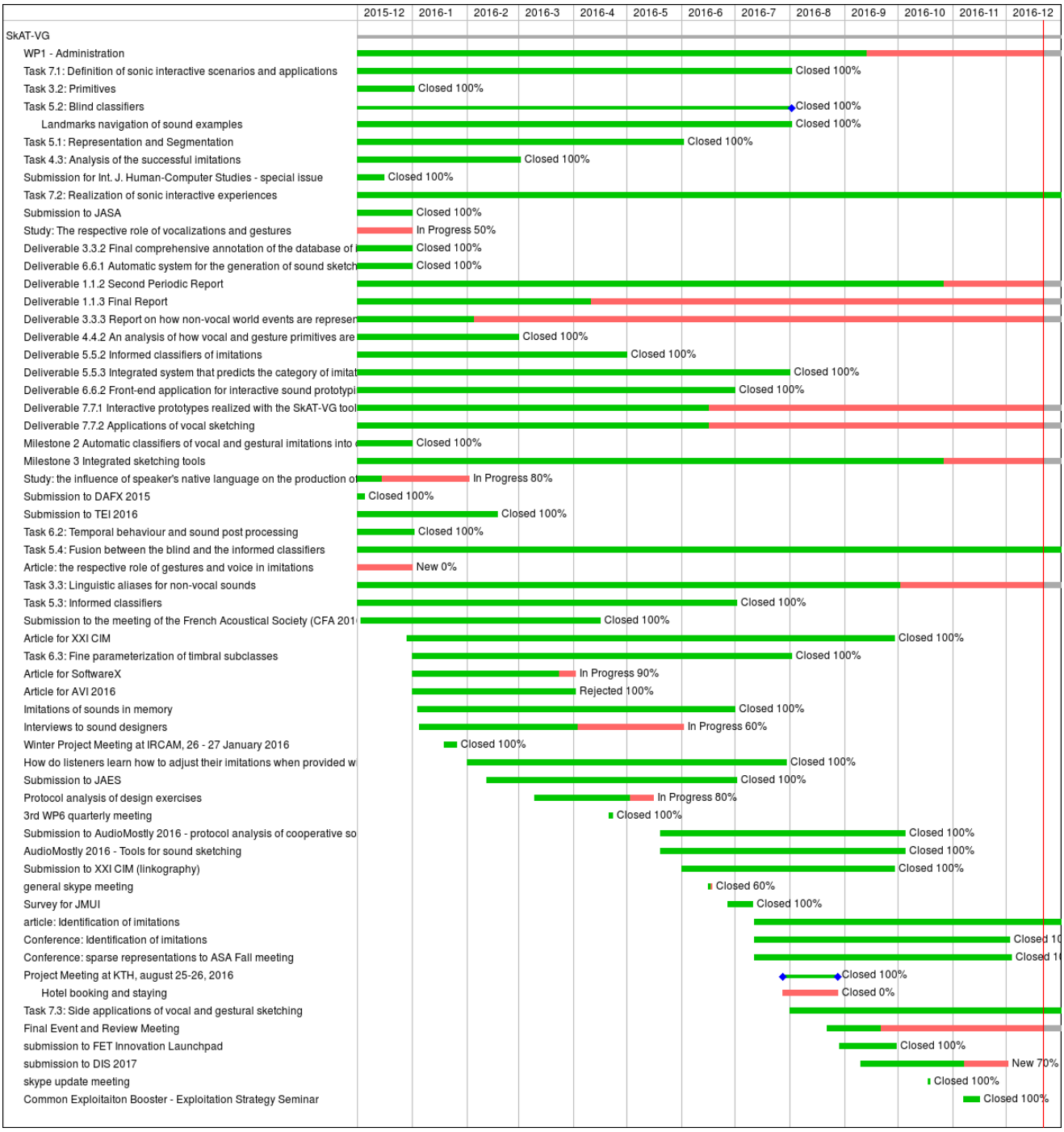


Figure 3: Activities of Year 3 as extracted from the Redmine project management tool on december 21, 2016.

WP	Description	Task	Progress	Deviation
WP1	Coordination		AtS	No
WP2	Case studies	T2.1	Completed	No
		T2.2	Completed	Minor slight delay
WP3	Production	T3.1	Completed	No
		T3.2	Completed	Yes discrete annotation, more PM
		T3.3	Completed	No
WP4	Perception	T4.1	Completed	No
		T4.2	Completed	Yes Statistical analysis of the whole database, annotation of only a subpart
		T4.3	Completed	No
WP5	Machine Learning	T5.1	Completed	No
		T5.2	Completed	No
		T5.3	Completed	No
		T5.4	Completed	No
WP6	Synthesis	T6.1	Completed	No
		T6.2	Completed	No
		T6.3	Completed	Minor anticipation
WP7	Design	T7.1	Completed	Minor anticipation
		T7.2	Completed	Yes prototype SEeD is not exploiting classification provided by T5.4
		T7.3	Completed	No

Table 2: Task progress and deviations.

Partner	WPs	Deviation
IUAV	WP6	Anticipation: prototypes for vocal and gestural sketching of T6.3
	WP7	Anticipation: automatic clustering, extension of sound synthesis palette
IRCAM	WP4	Modification: statistical analysis of the whole database, annotation of a subpart
KTH	WP2	Extension: additional subjects have been recorded in year 2
	WP3	Modification and Delay: discrete instead of time-continuous annotations; more PMs (transferred from WP2) needed for annotation
GENESIS	WP7	Anticipation: sound synthesis and control of vehicle sounds
IUAV IRCAM	WP7 WP5	The prototype SEeD is not exploiting the classifier provided by T5.4

Table 3: Deviations per partner.

2.3.1 Work Package 1: Project coordination

DoW: to ensure financial and administrative management of the project; to develop a spirit of co-operation between the partners; to ensure consensus management and information circulation among the partners, to ensure project reporting and interface with the Project Officer; to co-ordinate and control project activities to keep it within the objectives, to ensure quality management of the project.

Progress
Within WP1, resources are dedicated to manage the communication inside the project consortium with specific tools (e.g. Redmine, ownCloud) and towards the European Commission, to prepare and conduct project meetings and reviews, to prepare the minutes (see Section 2.4), to manage the fund transfers towards the partners, to monitor and report on the execution of the financial plan (more details in Section 4). A GANTT of the activities of Year 3, as extracted by the Redmine project management tool, is shown in Figure 3. The Project Management during the period is explained in section 2.4.

No Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

2.3.2 Work Package 2: Case studies

In WP2 the Consortium created an exploratory database of imitation case studies (Task 2.1). Then, a controlled procedure was devised for making a database of the productions of skilled imitators (Task 2.2).

Further details are available in the First Periodic Report.

Task 2.1: Case studies

DoW: Case studies will be collected from commercial recordings. Their quality will probably not be adequate, but they will be inspirational and reveal what skilled imitators can do. Then, a list of "action/sound primitives" will be defined: the classes of basic mechanical interactions that subjects can imitate. They will have to fulfill three requirements: being the simplest imitable sounding interactions, being combinable to form the sound events of Task 4.1 (IRCAM), and covering

the timbral families and scenarios of Tasks 6.1 and 7.1 (IUAV). The emphasis on "what the voice can do" imposes an approach based on the source filter model of sound production. Combining potential sources (trains of impulses, noises) and filters (from low- to high-Q resonators) a priori suggests a number of classes (impacts, whistles, bubbles, etc), but a precise set will be defined and limited at the output of Task 2.1.

Progress
Work on this task was described in the First Periodic Report.

No Deviations from Annex I
Task 2.1 is completed.

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 2.2: Set of imitations

DoW: High-quality recordings of imitations of the primitives will be made using skilled imitators, in a digital video format with both hi-fi audio (airborne and contact microphones), and video (frontal and profile views of mouth and hands). The video and contact microphone signals will assist the phonetic transcription in WP3. Task 2.2 will take care of using mechanical sounds, as well as sounds generated by IUAV in Task 6.1, as a the substrate of the imitations. This will ensure the compatibility with WP6 and WP7. Such classes will be analyzed one at a time in WP3.

Progress
Work on this task was described in the First Periodic Report.

Deviations from Annex I
For WP2, there are no deviations from Annex I, other than minor technical modifications of the setup, and a slight delay in the completion of recordings of skilled imitators.

All objectives achieved
The KTH recordings of skilled imitators were completed in early months of the second year of the project. The final annotation of these recordings was completed and used as input

for machine learning in WP5. A set of recordings of non-professionals made at IRCAM has also been annotated by the KTH team and added to the pool of data used for developing the informed articulatory classifiers in WP5.

No Corrective Actions Required

2.3.3 Work Package 3: Phonetic listening and world-event representation

This WP seeks to document how real-world events are imitated at the phonetic level. Recordings of skilled imitators were analyzed and annotated by phoneticians. The resulting labelling, paired with the audio, then forms the input data to the machine-learning developed in WP5.

Task 3.1: Editor

> **DoW:** Implementation of a data format and a graphical parameter editor for performing manual transcription of the selected articulatory parameters, from audio/video files to multitrack data files.

Progress
Work on this task was described in the First Periodic Report.

No deviations from Annex I
The task is completed.

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 3.2: Primitives

> **DoW:** Time-continuous phonological transcriptions of a number of imitations of the action primitives obtained in WP2. This will initially be done manually, using spectrograms, audio and video. These transcriptions will be in the form of continuous traces of manually estimated articulatory parameters (APs), connecting sequences of landmark points of recognizable phonetic configurations, i.e. phonemes. AP values represent the degree of activation of such phonetic sources as phonation, frication, plosives; and modifiers such as vowel resonances and stops.

AP values and phonemes will form the vocal primitives of the project. In unclear cases, the imitators will be subjected to direct articulatory measurements. The vocal primitives will be stored synchronously with the original audio, accompanied by annotations of gestural primitives, and processed by IRCAM in WP4 and WP5. For running a perceptual-feedback validation of the transcription work, it would be valuable, though not strictly necessary, to have an articulatory synthesizer that is driven to reproduce the vocal sounds interactively during the manual transcription. It is not clear that the current state-of-the-art in articulatory synthesis is good enough for this task. The existing TADA system from Haskins Laboratories will be tested as a possibility.

Progress
Work on this task was described in the Extra Periodic Report D1.1.1bis.

Deviations from Annex I
The choice of a system of discrete rather than time-continuous annotations is a deviation from Annex I. The annotations have been made not only of the action primitives, but also of more compound imitations.

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 3.3: Linguistic aliases for non-vocal sounds

DoW: When the target sound is very far from what is physically possible with the voice, sound symbolism in the form of onomatopoeias (invented sound-words) is what people generally use. KTH will consider how vocally inaccessible sounds might be specified using word-like aliases or semi-symbolic sounds. Because these are usually transcribed differently in different languages, SkAT-VG may initially need to define a new unambiguous phonetic representation, specific to the SkAT-VG system, that would have to be learned by its users. The International Phonetic Alphabet will be used as a starting point. By asking prospective SkAT-VG users to adopt a well-considered convention for such linguistic aliases, the system, being largely phoneme-based, could be trained or even constrained to map certain sound-words to non-vocal real-world events. A later refinement of the system (not in SkAT-VG) might ultimately make it more language-specific.

Progress
In 2015, a workshop was conducted in Copenhagen [ERDS16] in collaboration with WP7, to assess the propensity of sound designers to use linguistic aliases rather than actual imitations of sounds. The outcome indicated that a more stringent method of acquisition was called for, and that we needed to consider the interactive context, in particular how the type of recipient, human or non-human, affected the imitation strategy. Therefore in late 2016, a new experiment was performed. Given our experience from the Copenhagen workshop, clear examples of the use of tame imitations are the hardest to elicit, i.e., utterances that resemble true onomatopoetic words. We therefore designed the experiment so that it would maximise the elicitation of tame imitations and allow for wild ones. This was achieved by varying the interactive context from 1) a recipient who shared a language with the imitator, 2) to a recipient who did not share a language with the imitator, and finally 3) to a (Wizard-of-Oz) computer with audio input only. The experiment was conducted in the form of a collaborative game in which the imitators decribed referent sounds to recipients in the three conditions. The task of the imitator was to describe the sounds using any available communicative resource. This meant that the first condition maximised the linguistic embedding of the imitations, and the third condition minimised it. A total of 16 Swedish speaking imitators were recruited for the experiment and recorded on audio and video. The recordings were analysed and annotated using a score for the tame-wild continuum that was constructed with reference to similarity to speech. The scale considered the applicability of syllabicity, duration, voice quality, loudness, pitch and the use of native versus non-native speech segments. The results showed that, as hypothesised, the imitators were far more prone to produce tame imitations in condition 1 than in conditions 2 and 3. Thus, the linguistic context served to make the imitators produce imitations more closely resembling language, while interacting with a computer made them produce imitations that followed the acoustics of the referent sounds more closely. This observation indicates that tame input is not likely to be effective as an input channel for the SkAT-VG system. Another relevant finding was that the morphology of referent sounds strongly affected the imitation preferences on the tame-wild continuum. Referent sounds with a sound structure that resembles the structure of a syllable in speech are more likely to be imitated using speech like imitations. Referent sounds lacking such a structure do not easily lend themselves to be rendered through tame imitations. Hence, the inherent properties of the referent sounds influence the results.

Meetings and Events
The experiment here described was discussed at the general project meeting in Stockholm on august 2016.

No Deviations from Annex I
The Task is completed.

All Objectives achieved according to Schedule

No Corrective Actions Required

2.3.4 Work Package 4: Perception and cognition of vocalizations and expressive gestures

Overall, WP4 has three main objectives. First, WP4 aims at studying how people produce and perceive vocal and gestural imitations with *experimental studies*. The second objective is to provide the project (WP6 and WP7 in particular) with *datasets and new insights* on how vocal and gestural imitations can be practically used in the context of sound design. The third objective is to use vocal and gestural imitations as new tools to *investigate sound perception and cognition in general*. During the reporting period, WP4 mainly addressed the three objectives: First, it cleaned up, organized, described, and made the database of vocal imitations publicly available (it was previously only available to the consortium); Second, the consolidated analyses of the experiments carried out during the whole duration of the project has resulted in a better understanding of the phenomenon of vocal imitations in particular, and of sound perception and cognition in general. In terms of tasks, WP4 completed Tasks 4.1, 4.2, and 4.3. During the reporting period, WP4 has consolidated the results and published them in international journals (the Journal of the Acoustical Society of America and Plos One). We summarize here the main finding of these studies.

Task 4.1: Perceptually relevant categories of sounds

DoW: Perceptually discriminable categories of sounds will be defined. The task will be based on existing knowledge, and complement it by recording sounds, and conducting categorization and discrimination experiments to define different categories of sounds in terms of interaction, temporal and timbral properties. A strong interaction with IUAV will occur, to define the relevant temporal and timbral families of Task 6.1. Eventually, the outcome of Task 4.1 will consist of a large set of sounds, sorted in perceptually relevant categories. These categories will consist in combinations of the mechanical action primitives studied by KTH in Task 2.1.

Progress
Work on this task was described in the First Periodic Report.

No Deviations from Annex I
The Task is completed.

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 4.2: Successful imitations

DoW: Successful imitations will be sorted out by conducting identification experiments. Imitations of the perceptually relevant categories of sounds (Task 4.1) will be recorded focusing on different modalities: vocal and gesture. Particular attention will be drawn on expressive gestures related to temporal evolution of timbral properties based on the list of action/sound primitives defined in Task 2.1, and complementary characteristics found in Task 2.2. The outcome will be a large set of vocal and gestural imitations that successfully convey the different categories of sounds, and a set of gesture primitives in addition to the set of vocal primitives obtained in Task T2.2. This database will form the set of examples (vocal and gesture) required in Task 5.2. Task 4.2 will also study imitations of sounds generated by IUAVin Task 6.1. A specific methodology will be developed to handle the large number of sounds required by WP5: Tasks 4.1 and 4.2 therefore go in parallel with Tasks 2.1 and 2.2.

Progress

Identification of vocal imitations This study [LHV+17, LHV+16] investigated the semantic representations evoked by vocal imitations of sounds by experimentally quantifying how well listeners could match sounds to category labels. The experiment used three different types of sounds: recordings of easily identifiable sounds (sounds of human actions and manufactured products), human vocal imitations, and computational "auditory sketches" (created by algorithmic computations). The results showed that performance with the best vocal imitations was similar to the best auditory sketches for most categories of sounds, and even to the referent sounds themselves in some cases. More detailed analyses suggested that instead of trying to reproduce the referent sound as accurately as vocally possible, vocal imitations focus on a few important features, which depend on each particular sound category.

Analysis and publication of the data The Ircam database has been structured and restricted to complete data (audio, video and gesture data) in order to make them accessible to the scientific community. We also analyzed each imitation with acoustical descriptors developed within WP5 and added them to the available data. The two KTH databases (Deliverables D.3.3.1 and D.3.3.2) have been also cleaned and added to a common database. Now the database reflects the contributions of the different partners (IRCAM and KTH) and WPs (WP3, WP4 and WP5). In the back end, the repository is organized in three archives corresponding to each database. The front end, the website part, is in development and consists of a description of the database, the terms of use and the access with a login. The various copyright issues are still being settled. The database is going to be available to the public by the end of the project, and a companion journal paper is being jointly submitted.

Deviations from Annex I:
Instead of conducting identification experiments on the *whole* database of imitations, identification experiments have been conducted on a *sample* of the database of imitations, and statistical analyses have been conducted on the whole database.

All objectives achieved according the Schedule

No Corrective Actions Required

Task 4.3: Analysis of the successful imitations

DoW: Integration of the results of WP3 and WP4 aims at analyzing what makes an imitation successful. Identifying which sound features cannot be rendered by the human voice and gesture will inform Task 3.3 and Task 5.3 about the sounds that require linguistic aliases. The vocal and gesture primitives identified respectively by KTH in WP3 and IRCAM in Task 4.2 will be used to analyze the imitations, with a focus on the temporal combination. Multimodal analysis will inform on the pertinent gesture characteristics that can complement vocal imitation. The outcome of Task 4.3 will therefore inform two other tasks: 1. It will help WP5 refining its classifiers by providing it with correspondence rules between imitations and imitated sounds. 2. It will inform IUAV in Task 6.2 about the aspects that are important for the fine tuning of the timbral families.

Progress

Vocal imitations of basic auditory features The first of these studies [LJM+16, LJH+15] examined how vocal imitations of sounds enable their recognition by studying how two expert and two lay participants reproduced four basic auditory features: pitch, tempo, sharpness and onset. It used four sets of 16 referent sounds (modulated narrow-band noises and pure tones), based on one feature or crossing two of the four features. The four participants recorded vocal imitations of the four sets of sounds. Analyses identified three strategies: faithful reproduction, transposition, categorizations. Overall, these results highlight that vocal imitations do not simply mimic the referent sounds, but seek to emphasize the characteristic features of the referent sounds within the constraints of human vocal production.

Combining vocal and gestural imitations This study [SLF+15, SLF+16] had two parts. First (observational study), we manually annotated a subset of 80 combined vocal and gestural imitations, randomly drawn from the large database of imitations. The annotations focused on the pieces of information that were specific to gesture or vocalization. Analyses of the

co-occurences of vocal and gestural features suggested that imitators used different gestural metaphors to represent different sound features: imitators used a spatial metaphor to represent pitch (most of them associated pitch with the vertical dimension), produced pinch-like gestures to represent tonalness, and shaked their hands or arms to represent fluctuations in the sounds. The second part (experimental study) aimed to confirm these results by conducting an experiment with controlled synthetic stimuli, focusing on the spatial metaphor and shaking movements. Both manual annotations of the data and analysis of specifically developed gestural and vocal features confirmed and expanded the results. Lay imitators share common gestural metaphors: They represent pitch and spectral centroid along a vertical dimension, and shake their hands to represent fluctuations of the signals. Moreover, the regularity of the shaking gestures match the regularity of the sound fluctuations.

Using auditory bubbles to analyze vocal imitations At this stage of the project, we know that vocal imitations of everyday sounds are very well recognized. As such, they convey the pieces of information that are important for sound recognition. It is however not trivial to highlight what these features are. For instance, we have shown that vocal imitations reproduce faithfully certain acoustic features of the referent sounds, but also transform certain others to match the vocal ability of the imitators. To address this question, we used a method inspired by reverse correlation initially proposed in vision. The method consists of synthesizing sounds by randomly selecting "auditory bubbles" (small time-frequency glimpses) from the sounds' spectro-temporal representation, and then inverting the resulting sparsified representation. For each bubble selection, a decision procedure categorizes the resulting sound in one or the other category (voice or instrument). After hundreds of trials, the whole spectro-temporal space is explored, and adding together the correct answers reveals the relevant spectro-temporal patterns for each category. As a proof of concept [ISL16], we have collaborated with other researchers to apply this method to two categories of sounds for which we knew in advance what the characteristics features are: musical instruments (onsets) and voices (formants). The results highlighted higher frequencies (i.e. formants) for the voices, and lower frequencies (particularly during the onset) for the instruments, confirming the relevance of the method. Further work will apply the method to compare the important features in referent sounds and vocal imitations.

Meetings and Events
Quarterly meetings: January 27, 2016 (#7 together with project meeting in Paris); April 19, 2016 (#8 videoconferencing); August 28, 2016 (#9 together with project meeting in Stockholm).

No Deviations from Annex I

All objectives achieved according the Schedule

No Corrective Actions Required

2.3.5 Work Package 5: Automatic imitation recognition

WP5 deals with the automatic estimation of the categories of vocal imitations from the analysis of the audio and gesture signals. In this part we summarize the progress achieved in WP5 during the third year. Task 5.1 (representation and segmentation) and Task 5.2 (blind classifier of imitations) have been finalized during the second year (see Deliverable D5.5.1 at Mo28 for further details). Task 5.3 (informed classifier) and Task 5.4 (fusion classifier) have been finalized during the third year and are described here.

Task 5.1: Representation and Segmentation

DoW: The front-end application developed in WP5 will extract two types of meaningful representations of the imitation signals: Sound representations: low-level features (based on spectral moments, modulation spectrum, etc.) and perceptual features (loudness, pitch, sharpness, roughness, etc.); Gesture representations: gesture primitives obtained from WP4. These representations will be used to predict the high-level representations (vocal primitives) of WP3. This learning will be performed using the examples and analyses provided by WP3 and WP4. The mapping from sound to vocal primitives will be done by KTH. Imitations also involve the complex combination of vocal and gestural primitives. The front-end application of WP5 will segment the imitations into sequences of meaningful multimodal elements.

Progress
Progress on Task 5.1 has been extensively reported on the Extra Periodic Report D1.1.1bis and on the Deliverable D5.5.1 at Mo28.

Meetings and Events

No Deviations from Annex I
Task 5.1 is completed.

All objectives achieved according the Schedule

No Corrective Actions Required

Task 5.2: Blind classifiers

DoW: Classifiers that predict the categories of imitated sound sources (from WP4) directly from low-level multimodal (sound and gesture) representations.

Progress

Progress on Task 5.2 has been extensively reported on the Extra Periodic Report D1.1.1bis, on the Deliverable D5.5.1 at Mo28 and on the Deliverable D5.5.3 at Mo31. The classification system integrated in Max has been extended. First, the Dynamic Time Warping method described in D5.5.1 for the vocalization classification has been ported to Max as an external. Second, the gesture decomposition using Wavelet and Non-Negative Matrix Factorization (NMF) has also been ported to Max. These software modules can complete the current Max prototypes. In Year 3, KTH continued working on identification of vocal articulatory primitives from audio, and an extension to deliverable D5.5.1 was provided. Improvements have been introduced regarding the database, the annotation, the feature computation, and the machine learning methods.

Meetings and Events

No Deviations from Annex I

Task 5.2 is completed.

All objectives achieved according the Schedule

No Corrective Actions Required

Task 5.3: Informed classifiers

DoW: Classifiers that predict the categories of imitated sound sources (from WP4) from the high-level phonetic representations (from WP3) (instead of the low-level multimodal representations). Since high-level representations are used to model the categories, the statistical models will remain tractable by a human.

Progress
In this part we provide a summary of the work and results obtained in Task 5.3. Detailed information can be found in the Deliverable D5.5.2 at Mo28. The Task 5.3 "Informed Classifier" relates to the automatic recognition of the imitation categories starting from the transcription of an audio signal into vocal primitives (VPs). In D5.5.1, KTH has proposed a system to automatically transcribe an audio signal into these VPs. During the third year of the project, we studied the automatic recognition of the vocal imitation categories from this transcription. For each file, the KTH transcription system outputs: (T) the global phonetic transcription, (S) the global observation probability and (L) the set of global audio features used by the acoustic model. We used (T), (S) and (L) as input to a set of soft-margin SVM-RBF classifier to recognize the imitation categories. The best results were obtained using (L), i.e. the global audio features used for the transcription instead of the transcription itself. Moreover, these results remain lower than the ones obtained using the morphological audio features proposed in D5.5.1. Explanations for this can be either that 1) there is no relationship between vocal primitives and vocal/gesture imitation categories; 2) there is a relationship between vocal primitives and vocal/gesture imitation categories but the current KTH system failed to detect the vocal primitives on the vocal/gesture imitation dataset; 3) there is a relationship between vocal primitives and vocal/gesture imitation categories and the current KTH system performs correctly but the number of vocal primitives is not large enough to represent all the vocal mechanisms used in the vocal/gesture imitation dataset. Given the data we had, it was not possible to decide among these explanations. We then came back to the initial goal of task 5.3: develop a system to recognize the imitation categories inspired by speech recognition, i.e. including 1) a language model (a model that represents the set of possible sequences of phonemes used by people to imitate a given imitation category), 2) an acoustic model (a model that allows to recognize the phonemes over time from their acoustic occurrences). The idea was then to recognize the categories by decoding a set of hidden Markov models trained for each imitation category. Unfortunately, it was not possible to transcribe the vocal/gesture imitation dataset into vocal primitives. It was therefore not possible to train a language model. Also the system that transcribes an audio file into vocal primitives only performs at the file level (no transcription over time).

Method to automatically estimate the audio primitives. In order to overcome the aforementioned problems, IRCAM studied during this third year an innovative and promising methodology that allows to automatically estimate the definition of audio primitives, to automatically get their locations in time and use them in the framework of a language/acoustic model in the form of a set of hidden Markov models. In this approach, the primitives are not manually annotated but are automatically learned using unsupervised learning algorithm from a dataset of recordings. These primitives are named "audio primitives" since they do not rely on any vocal production mechanism but only on time and frequency audio representation. For this task, we propose the use of the Shift-Invariant Probabilistic Latent Component Analysis (SI-PLCA), the kernels of which are considered as "audio primitives". SI-PLCA was chosen since it allows to fulfill the two important requirements: additivity of the primitives (so that a given time-segment can be represented as the superposition of several primitives, for example a low-frequency harmonic component with super-imposed to it a high-frequency noise sound) and shift-invariance in time and frequency (the audio signal is represented by a reduced set

of primitives shifted in time and frequencies). Using a small annotated dataset, we show that such automatically derived audio primitives represent acoustically meaningful cues. We then develop a system that allows recognizing vocal imitations. For this the activations over time of the vocal primitives are considered as emissions of the hidden states of a Markov model. The succession of states is specific to each imitation category. We therefore represent each imitation category by a specific hidden Markov model.

Meetings and Events
The developments on informed classifiers were discussed in the following meetings:

- January, 26-27, 2016, Project Meeting, Paris, France
- August, 25-26, 2016; Project Meeting, Stockholm, Sweden

No Deviations from Annex I
Task 5.3 is completed.

All objectives achieved according the Schedule

No Corrective Actions Required

Task 5.4: Fusion

> **DoW:** Fusion between the blind and the informed classifiers. The performances of the two approaches (blind and informed) will be compared. From the results, fusion techniques will be developed to take advantage from both approaches (robustness versus precision). For each task, modules will be adapted in order to deal with the linguistic aliases (Task 3.3).

Progress
In this part we provide a summary of the work and results obtained in Task 5.4. Detailed information can be found in the Deliverable D5.5.2 at Mo28. The Task 5.4 "Fusion Classifier" relates to the automatic recognition of the imitation categories using the fusion (early or late) of the information provided by the Vocal Primitives transcription and the set of IRCAM audio features proposed in Deliverable D5.5.1. We tested the use of (T), (S) and (L) in complement to IRCAM morphological audio features to recognize the imitation categories. The best results were obtained using early-fusion of the (S) configuration and the morphological audio features. However, no benefits (except for the categories of the Machine family) were found compared to the use of only the morphological audio features.

Meetings and Events
The developments on fusion classifiers were discussed in the following meetings:

- January, 26-27, 2016, Project Meeting, Paris, France
- June, 16, 2016: Skype meeting with WP7
- August, 25-26, 2016; Project Meeting, Stockholm, Sweden

Deviations from Annex I
Task 5.4 is completed. As from the DoW, the studies on classification, culminating in Task 5.4, were supposed to inform the construction of a real-time classifier to be integrated in a tool for sound designers (SEeD, in Task 7.2). Such classifier has been implemented and is available for Max (see Progress in Task 5.2), and has been tested for model selection in the sound design tool. However, a user-trained classifier was preferred in SEeD, as a mismatch was found between how the database of sound categories was constructed and how the creative sound design process naturally unfolds. More precisely, the classifiers of WP5 were trained on imitations provided by participants who tried to reproduce a small number of reference sounds. On the other hand, users of the tool try to imitate a sound category from their internalized idea of such category, with no reference sound. Such deviation is attributed to WP7.

All objectives achieved according the Schedule

Corrective Actions
The DoW described a corrective measure to be taken if SkAT-VG would fail at "developing a user-independent system" (DoW, sec. 1.3.1, pag. 15). The deviation described above corresponds to such a corrective measure.

2.3.6 Work Package 6: Imitation-driven sound synthesis

In the first two years of the project, WP6 focused on two main tasks: Defining a set of relevant *timbral families*, namely providing sound synthesis tools to render the ontology of perceptually relevant sound categories defined in WP4 (T6.1), and calibrating these tools for expressive vocal control and temporal behaviour through effective mapping strategies between vocal features and synthesis parameters and sound post processing operations (T6.2). The achievement of these objectives is documented in D.6.6.1, "Automatic system for the generation of sound sketches", marked as P (prototype) and therefore including a preliminary release of the two pieces of software resulting from the accomplishment of the tasks: A major update of the Sound Design Toolkit (SDT), software package developed by IUAV providing a palette of physically informed sound synthesis models for education and research in Sonic Interaction Design, and SkAT-Studio, a modular framework developed by Genesis, designed to compose sound design workflows. In the second half of the project, WP6 has been mainly devoted to the development of prototype applications for imitation-driven sonic sketching (T6.3). Two

main products emerged from this activity: *miMic* and *Mimes*. The first is an augmented microphone with buttons and inertial sensors, enabling embodied sound design by automatically classifying vocal imitations into sound categories and subsequently rendering them by sound synthesis models controllable in real time by vocal and gestural input. The second is a system in which vocal imitations are used as source material, ready to be interactively manipulated by gestures applied to sensorized physical devices. The main idea behind both the products is to provide tools for sketching sound with voice and gesture, as immediate as a pencil is for sketching on paper. The goal is to let sound designers use their innate abilities, namely vocal and gestural production, to obtain rough but intuitive and immediate renditions of sonic concepts, leaving the refinement process to a later stage and to more conventional user interfaces. The design of the applications is based on the feedbacks collected through continuous exchanges with WP7 experimental activities, such as design workshops, interviews and case studies. The development ran concurrently with the studies of WP5, which informed the design choices on the classification modules. WP6 officially ended at Mo30 with Deliverable D6.6.2, "Front-end application for interactive prototyping", which is marked as R (Report) and thoroughly covers the design and development process of miMic and Mimes.

Task 6.1: Definition of the timbral families

DoW: Physics-based sound synthesis makes an extensive work of parameterization necessary, to define the different timbral families of each model. The specification of appropriate timbral families of the SkAT-VG system will be the main activity of this task. In particular, they will have to match the categories sorted out by WP3 and WP4, and will define the subclasses onto which the outputs of WP5 will be mapped.

Progress
Task 6.1 was completed before the end of the second year, and its progress has been already reported on the Extra Periodic Report D1.1.1bis.

No Deviations from Annex I
Task 6.1 is completed.

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 6.2: Temporal behaviour and sound post processing

DoW: Once the roughly parameterized synthesis algorithm will be selected, it will be necessary to establish (i) how it behaves in time, (ii) how and if the output sound must be post-processed and, (iii) if a sequence of subclasses must be used, how to pass from one timbre to the other. This task is essentially about calibration and control of time-varying parameterizations.

Progress
Task 6.2 was completed before the end of the second year, and its progress has been already reported on the Extra Periodic Report D1.1.1bis.

No Deviations from Annex I
Task 6.2 is completed.

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 6.3: Fine parameterization of timbral subclasses

DoW: From sketch to prototype. It is necessary to give the designer the possibility of processing and refining the sounds automatically generated on the basis of the results of Tasks 6.1 and 6.2. This task involves the definition and development of an interface that a designer can intuitively use for the exploration of a neighborhood of the default sound (the sketch). Such a sound will represent a sort of reference point (landmark) that the designer will be able to repeatedly play with, towards the definition of a prototype. In particular, the interface will allow the designer to vary a set of perceptually meaningful physics-based parameters and a set of controllers of the temporal behaviour of the sound. This task will be performed in cooperation between IUAV and GENESIS.

Progress
Work for Task 6.3 followed two separate but complementary paths. One of them researched the use of human voice as a tool for intuitive and expressive temporal control of synthesized sound, and the relationships between the dynamics of vocal production and the physical properties of the simulated sound events. The other mostly explored the use of gestures, enabling timbral manipulation of sounds (including vocal imitations) through physical manipulation of

tangible objects. Each of these two different directions led to the development of a prototype system, each of them providing a different but equally intuitive interface for sketching sonic interactions.

Sketching sound with voice: miMic Task 6.3 saw the development of a hardware and software application for embodied sound design called *miMic*. The physical device is based on a modified microphone, with embedded inertial sensors and buttons. Vocal imitations of referent sounds are automatically classified into sound categories, and subsequently rendered by sound synthesis models controllable in real time by vocal and gestural input. The application offers two modes of operation: *Select* and *Play*. In the *Select* step, vocal imitations produced by the user are classified into a subset of referent sound categories (8 out of the 26 defined in WP4). The output of the classification step is a mix of up to three timbral families, as implemented in the Sound Design Toolkit (outcome of Task 6.1 and documented in Deliverable D6.6.1), weighted according to the relevance of the corresponding sound category. In the *Play* step, vocal and gestural input is used to control in real time the synthesis parameters offered by the selected timbral families.

Analysis of the vocal signal Vocal control is achieved by analyzing the audio input coming from the microphone, extracting salient timbral descriptors as defined in Task 6.2 and then devising a strategy to map them to the control parameters offered by the synthesis models. One of the most important findings in the study of imitation-driven sound synthesis is that speakers are able to faithfully articulate only a limited number of timbral features at the same time, therefore limiting the vocal control of a given timbral family to a small subset of all the available audio descriptors. Another important observed fact is the recurrence of common strategies in the imitation of some timbral families, whose peculiarities are best captured by different sets of descriptors. For instance, onset detection conveys useful information about impulsive sounds, like *hitting* or *shooting*, while fundamental frequency estimation is more relevant in pitched sounds, like those made by electric motors and combustion engines. In Task 6.3, these observations informed the design and fine tuning of a smaller set of four higher-level audio descriptors, obtained by the aggregation of several mutually related low-level audio descriptors and roughly corresponding to the four vocal primitives defined in WP3: Phonation, myoelastic activity, turbulence and impulses.

Definition of the control layer The association between the audio descriptors extracted from the vocal/gestural signals (Task 6.2) and the synthesis parameters made available by the timbral families (Task 6.1) is defined by hand, striving to meet the expectations of the listeners about the sonic behavior of the synthesis models with respect to the imitations provided by the users. A manually defined mapping strategy based on the knowledge of an expert was preferred over other approaches, such as the use of machine learning techniques, because it is more consistent with the concept of ecological listening on which the whole project is built. In many cases, it is possible to exploit almost directly the common relations between timbral features and physical parameters. For instance, the *pitch* of a vocal signal can be directly mapped to the rate of a periodic process such as the revolutions of a combustion engine or an electric motor. As another example, the spectral *centroid* of an imitation can

be often related to the concept of size of resonating solids, bubbles in a liquid or objects struck by wind. Relations become even more evident with the new higher level descriptors: For instance, the rough and irregular texture of a palate grind, effectively captured by the *myoelastic* descriptor, immediately evokes rough and grinding sound events such as crumpling, scraping, rolling and so on. Although representing a marginal part of the control layer, the miMic hardware is equipped with an Inertial Measurement Unit (IMU) which allows to capture manual gestures, which often accompany vocal production to mimic the sound producing event or communicate morphological aspects of the sound. Signals coming from the IMU are mapped to sound synthesis parameters similarly to what happens for the audio features, to reinforce or complement the control actions exerted by the voice.

Improvement of the synthesis models Although timbral families were already fully defined in Task 6.1, some of the synthesis algorithms required modifications to be better controllable by vocal and gestural input. In Task 6.3, the algorithms describing impacts and frictions have been redesigned adopting an energy conservation scheme in the underlying physical models. This feature guarantees stability and convergence of the simulation for every possible combination of synthesis parameters, even the most extreme changes triggered by rapidly evolving vocal imitations.

Sketching sound with gestures: Mimes This prototype developed for Task 6.3 focuses on the use of gestures to control in real-time the manipulation and resynthesis of pre-recorded sounds, including vocal imitations. This application involves the use of a tangible interface composed of sensorized objects, and its sound generation system relies on *corpus-based concatenative synthesis* rather than physically informed models. The recorded sound material is divided in small fragments, which are analyzed in terms of timbral properties. Gestural signals coming from the sensors define trajectories in this timbral space, resulting in a non-linear recombination of the sound fragments. Although very different from miMic, Mimes follows a similar approach by implementing two modes of use, suitable for immediate exploration and later refinement of vocal sketches: 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. The recorded vocal imitation can be used as source material itself, or mapped to the fragments of another corpus based on their timbral description.

Corpus-based concatenative synthesis In the *Record* mode a vocal imitation is recorded in a buffer, 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 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.

Gestural control 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 maps can be obtained using the *mapping by demonstration* method. In this case, an example gesture is recorded and mapped to a vocal imitation by a hidden Markov model. 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.

Meetings and events

- WP6 third quarterly meeting (via Skype), 21st April 2016;
- Both miMic and Mimes have been used at the 48 hours of Sound Design Workshop at Château La Coste (27th april – May 1st, 2016);
- Demonstration of miMic for the World Voice Day at IUAV, Venice, 15th april 2016;
- Demonstration of miMic for the European Researchers' Night at IUAV, Venice, 30th September 2016.

No Deviations from Annex I
Task 6.3 is completed.

All Objectives achieved according to Schedule

No Corrective Actions Required

2.3.7 Work Package 7: Sonic Interaction Design

During the reporting period, WP7 research-through-design activities focused on a two-folded objective:

- Developing a fine-grained understanding of the creative processes enabled through the SkAT-VG methodology;
- Evaluating in context the effectiveness of the SkAT-VG tools in mediating and supporting creativity in sketching tasks.

The set of basic training activities, devised in the previous reporting period with the active collaboration with the other WPs, were refined and structured in a workshop format on vocal and gestural sketching. The emerging methodology set the experimental playground through which carrying out qualitative and quantitative assessment of the SkAT-VG project outcomes. Two main workshops on concrete sound design cases were organized: (i) the "48 hours of Sound Design at Château La Coste", and (ii) the industrial case study at PSA (French car manufacturer), which saw the active engagement of professional sound designers. Protocol and linkographic analyses, well-established approaches in visual design cognition studies, have been introduced for the first time in the sound design domain. The analyses provided relevant qualitative and quantitative information on the efficiency of the creative processes in cooperative vocal sketching tasks and, in this respect, for the further development of the SkAT-VG tools. The three demonstrators, SkAT Studio, miMic, and Mimes eventually blended in the software prototype SEeD (Sonic Embodied Design). This tool was actively used in the PSA industrial case study, and object of experimental assessment with professional sound designers.

Task 7.1: Definition of sonic interactive scenarios and applications

DoW: The aim will be to coordinate all the basic research efforts, with the definition and realization of concrete experiences that will serve as test benches of the SkAT-VG tool, as described in task 7.2. The task will consist of user studies and will be benefit form GENESIS experience in industrial sound. It will be active along most of the project duration. It will involve an intensive exchange with almost all of the other WPs and partners, and in particular WP2, WP3 and WP4 to distill relevant cases. The monitoring role of this task will be important to focus research toward applicable results, to be assessed in a number of design experiences.

Progress

Progress on Task 7.1 has been extensively reported on Extra Periodic Report D1.1.1bis. Here we only describe the third-year progress. Task 7.1 represents the SkAT-VG sketch book and sets the playground for the design exploration of the other WPs work. In other words, scientific contributions are nurtured from the design perspective, and mature sound design hypotheses are selected and investigated through workshops and mock-ups in Task 7.2. In the first reporting period, Genesis conducted several interviews with sound designers from France, Italy, UK and Japan, in order to get perspective on how vocal imitations may have a role or can impact the workflow of sound designers. In this third year, Genesis conducted three new interviews, and made a summary [DBM+17] of all interviews, thus providing further information to improve the SkAT-VG sketching tools. Genesis organized four new meetings, including one co-organized with IUAV.

Meetings and Events

- Julien Bayle came to Genesis office to train two employees on Ableton Live software and to present his workflow and methodology in sound design and sound creation. Bayle is a sound and visual artist, who creates programmed installations and audiovisual live performances. This meeting was also the occasion for Genesis to interview him;
- A full interview was done on Skype with Simon Cacheux, a sound designer who graduated at the French sound engineering school ENS Louis-Lumière. He worked a lot on sound scenography, and is now working for the automotive industry. He also teaches sound design at different schools;
- A full interview was done on Skype with Allister Sinclair, who is a sound designer mostly working for artistic projects, and has the particularity to use exclusively Pure Data to compose sounds. He studied at les Beaux Arts, Cergy, France;
- Hélène Lachambre and Clement Dendievel (GENESIS) visited IUAV from March 8th to 11th, 2016. The aim of this meeting was to plan the next workshops, and to set several specifications for the sketching tools that would be experimented during these workshops.

No Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

Task 7.2: Realization of sonic interactive experiences, using the SkAT-VG tool

DoW: Implementation by IUAV of sonically augmented mock-ups, selected from the applications emerging from Task 7.1. The definition and realization of real design sessions will aim at providing user tests of the capabilities of the SkAT-VG tool. In a series of workshops, designers will be invited to solve some design problems under a set of constraints. The analysis of the workshops, both in terms of reports and prototypes, will provide the assessment bench of the results of the project. The most compelling design processes will be documented with short movies, thus giving immediate evidence of the effectiveness of the SkAT-VG system. In order to face the risks discussed in Section v), the definition and the integration of auxiliary tools, based on other criteria than imitation (i.e. the linguistic aliases developed in Task 3.3), could be considered in the context of the design workshops.

Progress

During the third year, two events took place to make designers use the tools developed during the SkAT-VG project in a real world context: the "48 hours of sound design", an artistic event, and the "PSA case study", an industrial application. These two events are described in more details below, and extensively in Deliverable D7.7.2. In the third year, and thanks to a very close collaboration between IUAV and GENESIS, the first version of SkAT Studio came out. SkAT Studio, miMic and Mimes were tried out by five sound designers during the "48h of sound design" workshop. Thanks to the valuable feedback of the sound designers, IUAV, IRCAM and GENESIS made some reconsiderations about SkAT Studio, that have been functional to the development of a new tool: SEeD. GENESIS also developed the idea of a tool to draw sounds, that emerged from the interviews with sound designers. A new sound design module has been created in LEA (GENESIS software) in that sense. Both SEeD and the new LEA module were then experimented in an automotive industrial case-study with PSA.

Iuav and Ircam have collaborated to develop and evaluate SEeD. Three sound designers (who had participated to the 48 hours of sound design) as well as several sound design students participated to the evaluation. Their task consisted in using SEeD to re-create eight target sounds previously created with SEeD, in a very short time (4 minutes). The resulting sketches were compared to the targets by means of a distance metric developed in WP4 and described in D4.4.2. The results of the analysis showed that the sound designers were able to control SEeD intuitively, although not uniformly well for the whole repertory of models.

Meetings and Events

- 48 hours of Sound Design Workshop at Château La Coste: 27th april – May 1st, 2016. *The 48h of sound design* was a workshop organized by GENESIS in Château la Coste[2]. This workshop was inspired by the workshops previously organised by IUAV. In 48 hours, 5 sound designers (Andrea Cera, Allister Sinclair, Mathieu Pellerin, Simon Cacheux and Xavier Collet) tackled one artwork and proposed their aural interpretation of the selected piece using SkAT-VG tools and methods. The event started on April 28 and ended on May 1 2016 and had 4 main stages.

    1. *Introduction and training*. The SkAT-VG team introduced the overall workshop organization and SkAT-VG tools and methods to sound designers. The training session was about playing games and doing exercises on vocalizations, gestures, with and without SkAT-VG tools. During this training day, they also had to work collaboratively to sonify one video. This new collaborative way of creating sounds was really innovative and appreciated by the sound designers. At the end of the day, each sound designer selected an art piece for which they created a sound in the next 48 hours;
    2. *Sketching stage*. On April 29 morning, each sound designer met an art expert playing the role of a customer, to discuss the sound concept of the selected art piece, and to better define the details of their project. Following this meeting, they

[2]Château la Coste is a vineyard acquired in 2004 by Patrick McKillen to establish a center of both art and oenology. (`http://chateau-la-coste.com/en/`)

produced sketches all day long, by using exclusively SkAT-VG tools: SkAT Studio with Mimic sound models, and Mimes. By late afternoon, each designer discussed again with the customer about the sketches they produced in 24h, to guide the final sound composition. At this stage, the sound designers spontaneously decided not to use the tools, preferring to keep using vocalizations and gestures while discussing about sounds. This attitude was not registered in the early interviews at the beginning of the project.

3. *Refining stage and public presentation*. On April 30, the sound designers refined and expanded their sketches, by creating or adding new sounds. They were allowed to use both SkAT-VG tools (and some designers did) and their own tools. At the end of the day, more than 100 visitors had the opportunity to listen to sound designers' creation, on site in the Domain.
4. *Sound designers' feedback*. On April 30 and May 1, a second session of interviews were conducted to get sound designers' feedback on the use of sketches, on the use of voice and gestures, and on SkAT-VG tools. The sounds remained available for visitors on May 1.

The tools that were presented during this workshop did not exactly fit to the sound designers' needs, even though sound designers were nicely surprised of what they were able to do with the SkAT-VG tools in such a short time. The post-workshop reconsideration informed the creation of SEeD, that is combining some aspects of miMic and Mimes, within a new structure.

- PSA Industrial Case Study in Paris: 7th–8th November, 2016. This case-study was organised by GENESIS, and was the second concrete experience of sound design with SkAT-VG tools. Stefano Delle Monache (IUAV) and Clement Dendievel (GENESIS) led this case-study at PSA in Velizy, France. Five PSA employees participated to the study (one sound design project manager, two sound designers, and two audio-digital engineers). SkAT-VG tools (SEeD and the new LEA module) were tested by these five people, with the initial aim of designing HMI and driving-aid sounds (lane departure warning line, blind spot monitoring, turn signals, etc.). This first use-case in HMI and driving aid sounds did not work well, for the following possible reasons:

  1. PSA already designed a lot of driving-aid sounds, thus being quite tricky for them to start a new design from scratch and to re-think entirely what they did in the past. Moreover, usually their work is heavily constrained by specifications, and their creativity is limited when designing this type of sounds;
  2. At that point, SEeD was not allowing flexible pitch-based control, and this was perceived by the PSA team as a limit when they wanted to create melodies. The LEA module also showed some ergonomy problems. The PSA team spent a short time to familiarize with the tools, and therefore they needed assistance during use;
  3. Five people were working together. This is quite a large group for a creative session, where the different personalities may come to conflict.

  In the second day, the goal of the case-study was reset and the PSA team worked on electric car sound design. The goal of this second use-case was to design sounds for

an old Peugeot 205 and an old Citroen ZX, as if they were electric cars today. This new use-case was a lot more fruitful, especially during a collaborative session, where two groups (one of three people and one of two people) worked together on sketches. The collaborative aspect in sound design, brought by SkAT-VG methods, was really appreciated by the PSA team. It provided lots of creative interactions that were not present in PSA sound creation before. This use-case was crucial for SkAT-VG tools and methods, firstly because these have been experienced in the industrial field, and secondly because PSA participants had different technical backgrounds and positions (manager, engineer, sound designer). These two days gave us lots of new guidelines to refine the SkAT-VG tools.

- Clement Dendievel (GENESIS) and Stefano Delle Monache (IUAV) meeting at IRCAM: 9th November, 2016. This meeting was organised right after the automotive case-study with PSA, to organize the analysis of the PSA collaborative session. In particular, after the experience of using linkographic protocol analysis to study some sessions of the Copenhagen workshop in November 2015, the same method was adopted to analyze the PSA sessions.

Meetings and Events
June, 16, 2016: Skype meeting with WP5

Deviations from Annex I
Task 7.3 is completed. The sound design process developed and studied in WP7, which is based on SkAT-VG tools, was expected to make use of the classifiers provided by WP5. This was actually tried by embedding a real-time classifiers based on Dynamic Time Warping. However, this advanced experimentation revealed a fundamental misalignment: On the one hand sound categories were constructed and imitations were collected in WP4 based on a few reference sounds per category; On the other hand, in the creative process of sound design by vocal sketching, classes of sounds are evoked from internalized representations, with no accessible reference sounds.

All Objectives achieved according to Schedule

Corrective Actions
The DoW described a corrective measure to be taken if SkAT-VG would fail at "developing a user-independent system" (DoW, sec. 1.3.1, pag. 15). The fact that the prototype tool SEeD uses an individually trained classifier corresponds to such corrective measure of the DoW ("user-dependent variations can be handled by adaptive classifiers").

Task 7.3: Side applications of vocal and gestural sketching

DoW: The task, led by GENESIS will explore possible applications of SkAT-VG in areas such as sound effects for movies, real-time interaction in games, and

sound information retrieval. Although these areas are not in the central focus of the project, the partners have the skills and interests that allow testing the SkAT-VG findings and technologies in a wide range of applications.

Progress

Active Sound Design (ASD) for Electric Vehicles For a long time, GENESIS has been working on engine sound design. For this, the company developed a complete framework named ASD, which allows quick and ergonomic sound design, from current sound recording and analysis, to design and test of new engine sounds. Up to now, this framework was only available for combustion engines. Durint the SkAT-VG project, Genesis enhanced this tool to make it also available for electric motors: with the physical demonstrator (geneBOX), it is now possible to get data from an electric vehicle, and therefore to desgn the sound of an electric vehicle.

LEA Sound design module During the interviews, some sound designers expressed their will to be able to "draw" the time-frequency representation of a sound, for example using a tablet. Taking this demand into account, Genesis developed a new module in LEA to allow this kind of sound creation and manipulation, which can be seen as the synthesis counterpart of the auditory bubbles introduced in Task 4.3 to analyze vocal imitations. This module allows the user to draw a pattern in the time-frequency domain, and to apply a texture on it. This texture may be a single sine wave, or a noise, or a more complex harmonic-plus-noise model. Each added component can then be modified (redraw the pattern, change the texture, move/stretch the pattern in the time frequency domain). A second part of the module allows the user to design complex non-stationary harmonic-plus-noise models, and to quickly listen to them for validation.

Non-speech voice for sonic interaction At IUAV, a catalogue of research and applications of non-speech voice for sonic interaction has been compiled and published [PR16].

Meetings and Events
The LEA sound design module was tested during the PSA case study on November 7-8, 2016.

No Deviations from Annex I

All Objectives achieved according to Schedule

No Corrective Actions Required

2.4 Project Management during the period

As in Years 1 and 2, project management has been relatively easy in Year 3 as well, due to the clear complementarity of project partners in terms of areas of activity. The points of interaction and exchange between partners have been easy to define and finalized to achieve specific research goals.

Project management and coordination are described in WP1, which

> **DoW:** has functions of financial and administrative management. Within WP1, resources are dedicated to manage the communication inside the project consortium and towards the European Commission, to prepare and conduct project meetings and reviews, to prepare the minutes, to manage the fund transfers towards the partners, to monitor and report on the execution of the financial plan. Resources are also dedicated to quality control, to assure that the development process follows the quality rules for the project. The measurable success factors for all other Work Packages are monitored in WP1.

The objectives of WP1,

> **DoW:** To ensure financial and administrative management of the project. To develop a spirit of co-operation between the partners. To ensure consensus management and information circulation among the partners. To ensure project reporting and interface with the Project Officer. To co-ordinate and control project activities to keep it within the objectives. To ensure quality management of the project.

have been achieved. No deviations from the workplan for Y3 were deemed nor observed.

Management

The management structure described in Section 2.1.1 of the DoW, and implemented as reported in the First Periodic Report and in the Extra Periodic Report, has changed only slightly in Y3. Namely, Claudia Meoli joined the project as Administrative Assistant, and Davide Andrea Mauro left IUAV. Mauro's role in leading WP1 has been taken by the project coordinator. The roles are:

Coordination Team IUAV

Project Manager Davide Rocchesso

Scientific Assistant Stefano Delle Monache

Administrative Assistants Ilaria Rosa, Claudia Meoli

Project Committee :

IUAV Local Manager Davide Rocchesso

IRCAM Local Manager Patrick Susini

KTH Local Manager Sten Ternström

GENESIS Local Manager Patrick Boussard

Work Package Leaders :

WP1 - IUAV Davide Rocchesso

WP2 - KTH Sten Ternström

WP3 - KTH Pétur Helgason

WP4 - IRCAM Guillaume Lemaitre

WP5 - IRCAM Geoffrey Peeters

WP6 - IUAV Stefano Baldan

WP7 - GENESIS Patrick Boussard

Quality assurance and risk management

According to what is specified in Section 1.3.1 of the DoW, and whenever it was possible, preventive actions were conceived at the project design stage to reduce the identified failure risks related to each WP. Contingency plans were activated at the occurrance of a risk. For these reasons, some works have been anticipated and three deviations have been registered, as reported in table 2.

Collaboration infrastructure and access to documents

The following elements of a collaboration infrastructure have been established since the beginning of the project:

Web sites and Social Networks `http://skatvg.iuav.it/` mirrored on `http://skatvg.eu/`, `https://twitter.com/SkATVG`, `https://vimeo.com/skatvg`. The website is intended as a showcase for the project itself. It encompasses a section for presenting the project, one section for presenting the partners involved in the project, and a number of sections to keep track of what is happening in the context of the project. Recent news and updates are "tweeted" to reach a wider audience.

Mailing lists General list `skat-vg@ircam.fr` used extensively as an efficient discussion platform and for general organizational purposes. Managed at IRCAM;

Coordinator list `skat-vg@iuav.it` used for internal purposes by the Coordination Team. Managed at IUAV.

Redmine project management tool `https://redmine.skatvg.iuav.it` Redmine is a flexible project management web application, which provides several useful features:

Issues This is the core functionality of the application. Everything regarded as important for the project can be raised as an "issue" (with various types and definitions) and the evolution can be managed, assigned to specific people, and monitored.

Documents and Files These repositories contain documents that can be shared between partners and be directly linked in the Issues.

Wiki A Wiki is used as a collaborative workplace to share information, e.g. Lists of similar and relevant technologies, Relevant Literature, Calls for Conferences, and so on.

Others (GANTT, Calendar, Activity) These facilities are used for specific purposes such as automatically producing GANTT charts (see Figure 3) sharing events on a calendar, keeping track of the overall project activity.

SVN repository `https://skatvg.iuav.it/svn/skatvg_svn`: A revision control system for collaboratively writing documents, publications, and reports (especially in LaTeX) and for collaborative software development. The repository is readable from the Redmine page and it exposes the directories: • Budget • Code • Data • Documents • Misc

GitHub repository `https://github.com/SkAT-VG`: GitHub is a Web-based Git repository hosting service, which offers all of the distributed revision control and source code management. It is the main public resource where to find the software outcomes of the project. It allows the researchers from the consortium to update the code while permitting to the general public to download the source code.

Build in Progress `http://buildinprogress.media.mit.edu` has been chosen to document the design and development process of some SkAT-VG prototypes.

Teleconferencing A number of solutions have been evaluated, with a decision to use Skype. Conference calling is an effective tool to reach consensus about technical issues among multiple partners. Minutes are kept in the project's Redmine installation.

Starting from the beginning, the project had planned intra- and inter-WPs group calls, where members updated each other on the development status of the individual WPs and on plans for the successive periods. In most cases more focused follow-up discussions were held by some of the partners.

Meetings and exchanges

Since January 2016, in its third year the project arranged the following meetings and research visits:

General Meetings

- Paris Project and Review Meeting: 26–27th January, 2016. Minutes of the meeting are available on Redmine;
- Stockholm Project Meeting Mo30: 25–26th August, 2016. Minutes of the meeting are available on Redmine;

- Common Exploitation Booster – Exploitation Strategy Seminar in Venice Mo32: 15th November, 2016. The Final Report is available in Redmine.

Cross-partner Meetings

- 48 hours of Sound Design at Château La Coste: 27th april – May 1st, 2016. Documentation available on Redmine and project web site;
- PSA Industrial Case Study in Paris: 7th–8th November, 2016. Documentation available on Redmine.

Teleconferences:

- WP4: Quartely meeting, 19th April, 2016;
- WP5, WP7: 16th June 2016;
- WP6: Quartely meeting, 21st April 2016;

Minutes from these meetings are available on Redmine.

Research Visits:

- Hélène Lachambre and Clement Dendievel (Genesis) at IUAV: 8-11th March, 2016.
- Clement Dendievel (Genesis) and Stefano Delle Monache (IUAV) at IRCAM: 9th November, 2016.

3 Deliverables and Milestones tables

3.1 Deliverables

All SkAT-VG deliverables (excluding the periodic reports) are summarized in Table 4. The year 3 deliverables are described as follows:

D3.3.3 Report on how non-vocal world events are represented phonetically

D4.4.2 An analysis of how vocal and gesture primitives are sequenced.

D5.5.2 Informed classifiers of imitations.

D5.5.3 Integrated system that predicts the category of imitated sound sources.

D6.6.2 Front-end application for interactive sound prototyping.

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

D7.7.2 Applications of vocal sketching.

3.2 Milestones

Milestone M1, "Accumulation of a large enough database of recorded, sorted, and labeled imitations", was achieved after Year 1.
Milestone M2, "Automatic classifiers of vocal and gestural imitations into categories of imitated sounds", was achieved after Year 2.
In Year 3, IRCAM has continued working to assess the success of vocal and gestural imitation in conveying the imitated referent sounds, to analyse the large database of imitations, and to develop classifiers and descriptors; KTH has extended the annotation work, including IRCAM recordings, has improved its articulatory classifiers, and has performed an experiment to assess the relevance of linguistic aliases in sound communication; The work of GENESIS centered around interactions with several sound design professionals and development of tools (SkAT Studio, LEA module) that address some of the needs of such stakeholders; IUAV has been further developing sound synthesis modules. IUAV, IRCAM, and GENESIS have been collaborating at the development of prototype applications. IUAV and GENESIS have been collaborating at the organization and analysis of design sessions and workshops (artistic workshop and industrial case study). Ten people external to the consortium (sound designers, audio digital engineers, manager) had extensive experimentation with SkAT-VG tools.
Milestone M3, "Integrated sketching tools", can be considered to be achieved at the end of Year 3.

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Dissem-ination level	Delivery date from Annex 1 (Mo)	Delivered Yes/No	Actual / Forecast delivery date (Mo)	Comments
2.2.1	Explorative collection of imitated sounds	2	KTH	R	PU	4	Yes	5	
2.2.2	Extensive set of recorded imitations	2	KTH	R	PU	12	Yes	12	
3.3.1	Preliminary annotation of the database of imitations of action primitives in terms of vocal primitives	3	KTH	R	PU	12	Yes	12 / 14	draft / update
3.3.2	Final comprehensive annotation of the database of imitations	3	KTH	R	PU	24	Yes	22 / 24	draft / update
3.3.3	Report on how non-vocal world events are represented phonetically in SkAT-VG	3	KTH	R	PU	36	Yes	36	
4.4.1	A large set of vocal and gestural imitations	4	IRCAM	R	PU	21	Yes	21	
4.4.2	An analysis of how vocal and gesture primitives are sequenced	4	IRCAM	R	PU	26	Yes	26	
5.5.1	Blind classifiers of imitations	5	IRCAM	R	PU	23	Yes	22 / 23 / 36	draft / update / extension
5.5.2	Informed classifiers of imitations	5	IRCAM	R	PU	28	Yes	28	
5.5.3	Integrated system that predicts the category of imitated sound sources	5	IRCAM	P	PU	31	Yes	31	
6.6.1	Automatic system for the generation of sound sketches	6	IUAV	P	PU	24	Yes	22 / 24	draft / update
6.6.2	Front-end application for interactive sound prototyping	6	IUAV	R	PU	30	Yes	30	
7.7.1	Interactive prototypes realized with the SkAT-VG tool	7	GENESIS	P	PU	36	Yes	36	
7.7.2	Applications of vocal sketching	7	GENESIS	R	PU	36	Yes	36	

Table 4: Deliverables of SkAT-VG in the three years. The approved deliverables, which are publicly available from the SkAT-VG website, are grayed-out in the table.

Milestone no.	Milestone name	WP no.	Lead beneficiary	Delivery date from Annex 1 (Mo)	Delivered Yes/No	Actual / Forecast delivery date (Mo)	Comments
M1	Accumulation of a large enough database of recorded, sorted, and labeled imitations	2, 3, 4	KTH	12	Yes	12/15	draft/update (Figure 2)
M2	Automatic classifiers of vocal and gestural imitations into categories of imitated sounds	3, 4, 5, 6	IRCAM	24	Yes	22/25	draft/update (Figure 2)
M3	Integrated sketching tools	3, 4, 5, 6, 7	IUAV	36	Yes	36/37	draft/update (Figure 2)

Table 5: Milestones of SkAT-VG in the three years. The milestones achieved in previous reporting periods are grayed out.

4 Explanation of the Use of Resources and Financial Statements

The following tables detail the major costs that occurred in the third year, in a form consistent with Section 2.4 of the DoW. The figures are approximate, as they have been computed in November 2016. Overall, the **estimated costs** are consistent with the planned budget.

Description	Resources	Cost
	IUAV	
Personnel	Research positions BALDAN S., researcher, 01/01/16 - ; DELLE MONACHE S., assistant professor, 01/01/16 -; CERA A., sound designer, 01/01/16 - ; ROCCHESSO D., professor, 01/01/16 -)	73174
	Management MEOLI C., administration collaborator 01/04/16 ROSA I., administration collaborator, 01/01/16 - ; ROCCHESSO I.	29777
Equipment	Computers	4648
	Software licenses	970
	Audiovisual equipment	2226
Travel	10 Conference trips DELLE MONACHE S.: Eindhoven (Netherlands) 13-18.02.2016; ROCCHESSO D.: Eindhoven (Netherlands) 13-18.02.2016; ROCCHESSO D.: Hamburg (Germany) 31.08-02.09.2016; DELLE MONACHE S.: Cagliari (Italy)– XXI CIM 27.09-01.10.2016; DELLE MONACHE S.: Norrkoping (Sweden) – Audiomostly 3-6.10.2016	5321
	Participations to 2 Project Meetings ROCCHESSO D.: Bruxelles+Paris 25-29.01.2016; DELLE MONACHE S.: Paris del 24-29.01.2016; BALDAN S.: Paris del 25-27.01.2016; ROCCHESSO D.: Stockholm (Sweden) 23-30.08.2016; BALDAN S.:Stockholm (Sweden) 23-30.08.2016; MEOLI C.: Stockholm (Sweden) 23-30.08.2016; DELLE MONACHE S.: Stockholm (Sweden) 24-27.08.2016	4378
	4 Participations to Review Meeting Final Meeting - Paris 19-20.01.2017	2000
	2 Research workshops DELLE MONACHE S.: Aix en Provence (France) 27.04-01.05.2016; BALDAN S.: Aix en Provence (France) 27.04-01.05.2016; ROCCHESSO D.: Paris-Aix en Provence (France) 26.04-01.05.2016; CERA A., Aix en Provence (France) del 7.04-01.05.2016; DELLE MONACHE S., Paris (France) 5-9.11.2016	2181
Dissemination	Publications and promotion	4100
Total direct costs		128775

Description	Resources	Cost
	IRCAM	
Personnel	Research positions (G. Lemaitre, V. Isnard, E. Marchetto, G. Meseguer,)	130129
	Permanent research and management positions (F. Bevilacqua, O. Houix, N. Misdariis, G. Peeters, P. Susini)	69621
Equipment	Computers	1648
	Audiovisual	6850
Other	Conference Registrations	2890
	Scientific Societies Registration	266
	Conference Posters	76
	Audit	1240
Travel	Participations to Project Meetings	6149
	Conference trip	3861
Total direct costs		222730

Description	Resources	Cost
	KTH	
Personnel	Research positions (Sten Ternström, Anders Friberg, Glaucia Salomão, Pétur Helgason, Tony Lindeberg, Tino Weinkauf, Gregorio Palmas, Christine Ericsdotter Nordgren)	161716
	Management (Sten Ternström)	21929
Other costs		7827
Travel	Meetings Paris, 01/16; Stockholm, 08/16; Final review Paris, 01/17)	10192
Total direct costs		201664

Description	Resources	Cost
	GENESIS	
Personnel	Research positions Patrick Boussard, researcher, 01/10/15 - ; Hélène Lachambre, researcher, 01/10/15 - ; Guillaume Stempfel, researcher, 01/10/15 - ; Stéphane Molla, 01/10/15 - ; Clément Dendievel, 29/02/16 -	90573
	Management Patrick Boussard, 01/10/15 -	5787
Equipment	Computers	1654
	Software licenses	499
	Audiovisual equipment	2566
	Documentation	2550
Travel	Participations to Project Meetings P. Boussard: Project/Review Meeting (Lisbon, Portugal), 10/15; P. Boussard, H. Lachambre, G. Stempfel, C. Dendievel: Project Meeting (Paris, France), 01/16; P. Boussard, C. Dendievel: Project meeting (Stockholm, Sweden), 08/16; P. Boussard, H. Lachambre, C. Dendievel: Project/Review Meeting (Paris, France), 01/17	6131
	Cross-partner meetings H. Lachambre, C. Dendievel: IUAV-Genesis (Venice, Italy), 03/16; C. Dendievel: IUAV-Genesis-IRCAM (Venice, Italy), 11/16	1893
	Conferences C. Dendievel, P. Boussard: Internoise (Hamburg, Germany), 08/16	1554
	SkAT-VG Events"48h of sound design" (Chateau Lacoste, France), 04/16; PSA case study (Paris, France), 11/16	4638
Total direct cost	Note: In D1.1.1bis, the Genesis explanation of the use of resources and financial statements was wrong, as it was taking into account 60% functioning cost. The total should have been 71742 instead of 114787.	117845

Table 6 and Figure 4 report the overview of Person-Months status (cumulative). Overall, the table gives account of a homogeneous development of the scientific work. Among the workpackages, the largest number of person months has been dedicated to WP4, one of the scientific cores of the project, and in turn this resulted in a considerable number of scientific publications in international journals.

PM Claimed	WP1	WP2	WP3	WP4	WP5	WP6	WP7	TOTAL
IUAV	10.66/12	3.38/4	3.38/3	4.29/3	2.95/2	35.07/30	19.02/15	78.75/69
IRCAM	3/3	2/2	4/4	63.31/48	28/28	4/4	4/4	108.31/93
KTH	3/3	10.5/16	38.2/32	0/2	8.4/8	0/3	0/4	60.1/68
GENESIS	2/4	6/6	0/0	6/6	6/6	11/8	25/20	56/50
Total	18.66/22	21.88/28	45.58/39	73.6/59	45.35/44	50.07/45	48.02/43	303.16/280

Table 6: Effort in Person-Months, per Workpackage and per partner. Whole duration of the project.

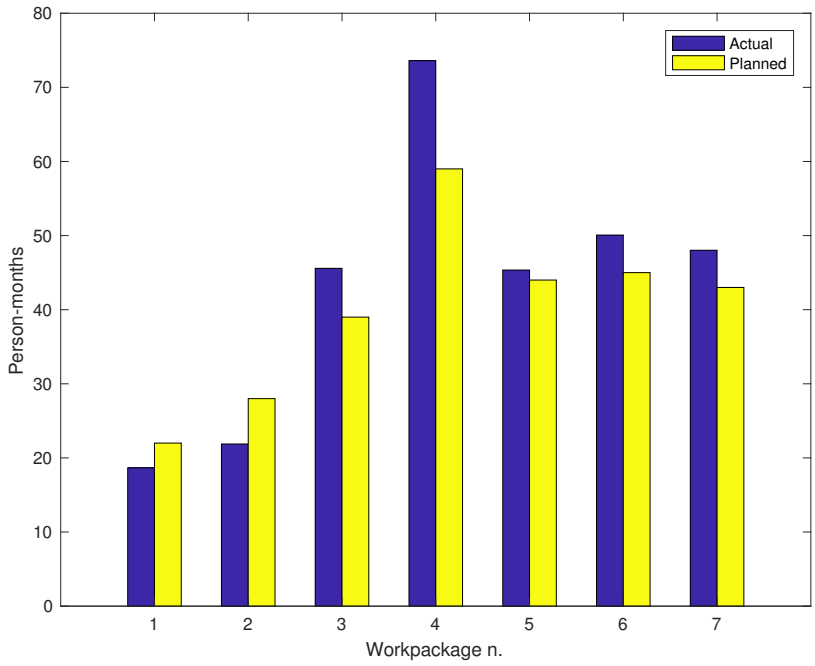


Figure 4: Person months of the project as compared to the overall planned effort over three years.

Table 7 and Figure 5 report the overview of Personnel and Other direct costs. Overall, the costs are slightly lower than expected.

Costs	Personnel	Other	TOTAL
IUAV	358738/394800	83701/117000	442439/511800
IRCAM	520972/487753	40744/91200	561716/578953
KTH	412797/445600	33958/80000	446755/525600
Genesis	271468/264550	46148/32666	317616/297216
Total	1563975/1592703	204551/320866	1768526/1913569

Table 7: Personnel and Other Direct Costs, per partner: Actual / Planned.

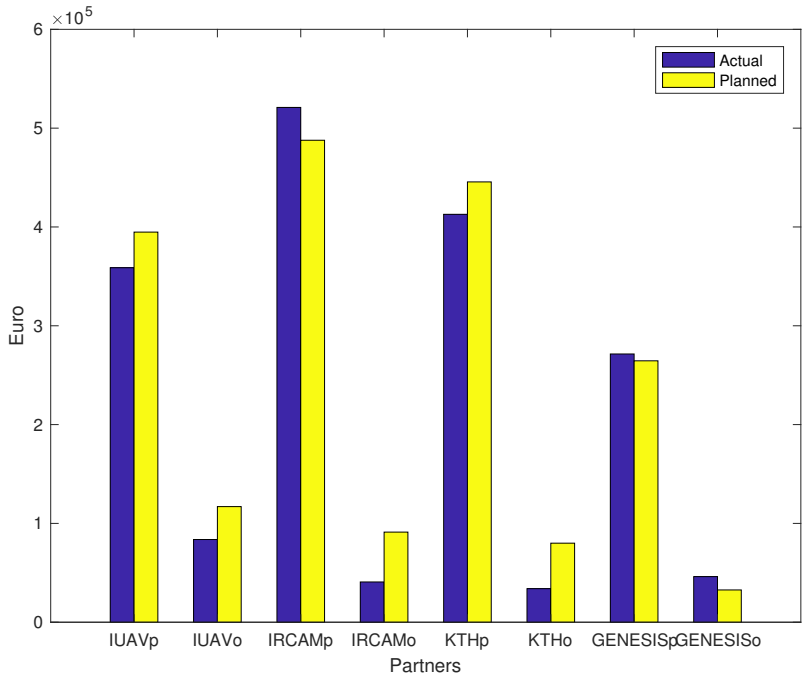


Figure 5: Costs per partner as compared to the overall planned costs over three years. "p": Personnel, "o": Other Direct Costs.

5 List of Publications, Networking, and Dissemination Activities

Publications:

1. "Imitating sounds with gestures", G. Lemaitre, H. Scurto, G., J. Françoise, P. Susini, F. Bevilacqua. Journal paper in preparation, 2017.

2. "Understanding and modifying sound design practices", C. Dendievel, S. Baldan, S. Delle Monache, H. Lachambre, D. Rocchesso, and P. Boussard. 2017. Journal paper in preparation.

3. "The sound design toolkit", S. Baldan, S. Delle Monache, and D. Rocchesso. SoftwareX (in press), 2017.

4. "Vocal imitations of non-vocal sounds." G. Lemaitre, O. Houix, F. Voisin, N. Misdariis, and P. Susini. Plos ONE 11(12): e0168167. DOI:10.1371/journal.pone.0168167

5. "Non-speech voice for sonic interaction: a catalogue", A. Del Piccolo and D. Rocchesso, J. Multimodal User Interfaces, 2016. DOI:10.1007/s12193-016-0227-6

6. "Organizing a Sonic Space through Vocal Imitations", D. Rocchesso, D.A. Mauro, C. Drioli. Journal of the Audio Engineering Society, 64(7/8):474-483, 2016.

7. "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):290–300, 2016.

8. "Multisensory texture exploration at the tip of the pen" D. Rocchesso, S. Delle Monache and S. Papetti. International Journal of Human-Computer Studies, vol. 85, pp. 47-56, 2016.

9. "Sketching sound with voice and gesture", D. Rocchesso, G. Lemaitre, P. Susini, S. Ternström, and P. Boussard. interactions 22:1, 2015, pp. 38-41.

10. "Idealized computational models for auditory receptive fields", T. Lindeberg and A. Friberg. PloS one, 10(3), 2015.

11. "Scale-space theory for auditory signals", T. Lindeberg and A. Friberg. In Scale Space and Variational Methods in Computer Vision, pages 3–15. Springer, 2015.

12. "Bauhaus legacy in Research through Design: the case of Basic Sonic Interaction Design." S. Delle Monache, D. Rocchesso. International Journal of Design [Online] 8:3, 28 December 2014.

13. "Sounding objects in Europe", D. Rocchesso. The New Soundtrack, Volume 4, Issue 2, pp 157–164, September, 2014.

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

15. "Might as Well Jump: Sound Affects Muscle Activation in Skateboarding" P. Cesari, I. Camponogara, S. Papetti, D. Rocchesso, and F. Fontana. PLOS ONE, Volume 9, Issue 3, 9 March 2014.

16. "Embodied sound design", D. Rocchesso, S. Baldan, F. Bevilacqua, A. Cera, S. Delle Monache, and G. Lemaitre. Paper submitted to conference. 2017.

17. "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, 140(4): 3390, 2016. Proceedings of the meeting of Acoustical Society of America, Honolulu, HI.

18. "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, 140(4): 3267, 2016. Proceedings of the meeting of Acoustical Society of America, Honolulu, HI.

19. "Cooperative sound design: A protocol analysis". S. Delle Monache and D. Rocchesso. In Proceedings of the Audio Mostly 2016 (AM '16). ACM, New York, NY, USA, 154-161, 2016.

20. "Innovative Tools for Sound Sketching Combining Vocalizations and Gestures". O. Houix, S. Delle Monache, H. Lachambre, F. Bevilacqua, D. Rocchesso, and G. Lemaitre. In Proceedings of the Audio Mostly 2016 (AM '16). ACM, New York, NY, USA, 12-19, 2016

21. "Sonic in(tro)spection by vocal sketching", A. Cera, D. Andrea Mauro, and D. Rocchesso. In Proceedings of the XXI Colloquio di Informatica Musicale, Cagliari, Italy, September 2016.

22. "Understanding cooperative sound design through linkographic analysis", S. Delle Monache and D. Rocchesso. In Proceedings of the XXI CIM Colloquium on Music Informatics, Cagliari, Italy, September 2016.

23. "Sketching sonic interactions by imitation-driven sound synthesis", S. Baldan, S. Delle Monache, D. Rocchesso, and H. Lachambre. In Proc. 13th Sound and Music Computing Conference, SMC 2016, pages 47–54, 2016.

24. "Sketching step in sound design: the sound designers" point of view", P. Boussard, C. Dendievel and H. Lachambre. In Proc. of Internoise, Hamburg, Germany, pages 4505-4513, August 2016.

25. "Embodying sounds: Building and analysis of a database of gestural and vocal imitations." H. Scurto, G. Lemaitre, J. Franọise, F. Bevilacqua, P. Susini, and F. Voisin. In Gesture, Creativity, Multimodality. Procedings of ISGS 7 : 7th Conference of the

International Society for Gesture Studies, Paris., page 269, The University of Texas at Austin, Austin, TX, 2016. International Society for Gesture Studies.

26. "A database of articulatory annotations of vocal imitations", P. Helgason, G. Laís Salomão, and S. Ternström, Proceedings of the XXIX Swedish Phonetics Conference, Stockholm (Sweden), 13 – 15 June, 2016.

27. "miMic: The microphone as a pencil" D. Rocchesso, D. Mauro, and S. Delle Monache. Proc. of the 10th International Conference on Tangible, Embedded and Embodied Interaction (TEI), Eindhoven (The Netherlands), 14-17 February 2016.

28. "A set of audio features for the morphological description of vocal imitations" E. Marchetto and G. Peeters. Digital Audio Effects Conference, Trondheim, Norway, 2015.

29. "Reverberation still in business: Thickening and propagating micro-textures in physics-based sound modeling" D. Rocchesso, S. Baldan, and S. Delle Monache. Digital Audio Effects Conference, Trondheim, Norway, 2015.

30. "Vocal imitations of basic auditory features" G. Lemaitre, A. Jabbari, O.Houix, N. Misdariis, P. Susini, The Journal of the Acoustical Society of America, vol. 137 (4), p. 2268, 2015. (Proceedings of the meeting of Acoustical Society of America, Pittsburgh, PA).

31. "Combining gestures and vocalizations to imitate sounds." H. Scurto, G. Lemaitre, J. Franọise, F. Voisin, F. Bevilacqua, and P. Susini. The Journal of the Acoustical Society of America, 138(3):1780–1780, 2015. Proceedings of the meeting of the Acoustical Society of America, Jacksonville, FL.

32. "Analyzing and organizing the sonic space of vocal imitation" D.A Mauro, and D. Rocchesso. Audio Mostly 2015, Thessaloniki (Greece), 07-09 October, 2015.

33. "To "Sketch a Scratch" " A. Del Piccolo, S. Delle Monache, D. Rocchesso, S. Papetti, D.A. Mauro. 12th Sound & Music Computing conference (SMC), Maynooth (Ireland), July 26 - August 01, 2015.

34. "Growing the practice of vocal sketching" S. Delle Monache, D. Rocchesso, S. Baldan, D.A. Mauro. 21st International Conference on Auditory Display (ICAD–2015), Graz (Austria), 07-10 July, 2015.

35. "Advanced signal processing methods for the analysis of transient radiated noise from submarines" T. Leissing, C. Audoly, H. Lachambre and G. Stempfel. Proc. of Internoise, Melbourne (Australia), 16-19 November, 2014

36. "Physically informed car engine sound synthesis for virtual and augmented environments", S. Baldan, H. Lachambre, S. Delle Monache, and P. Boussard. In Proceedings of the 2nd Workshop on Sonic Interactions in Virtual Environments - IEEE VR 2015, Arles, France, 2015.

37. "Non-Verbal Imitations as a Sketching Tool for Sound Design." G. Lemaitre, P. Susini, D. Rocchesso, C. Lambourg, P. Boussard, In Mitsuko Aramaki et al., editors, Lecture Notes in Computer Sciences : Sound, Music, and Motion. Springer, Berlin/Heidelberg, Germany, 2014, pp. 558-574.

38. "Self-organizing the space of vocal imitations" D. Rocchesso, and D.A. Mauro. XX Colloquio di Informatica Musicale, Rome (Italy), 20-22 October, 2014.

39. "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), 20-22 October, 2014.

40. "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), 14-20 September, 2014.

41. "2001-2016: Oggetti sonanti in Europa." D. Rocchesso, Associazione Italiana di Acustica 41° Convegno Nazionale, Pisa (Italy), 17-19 June, 2014.

42. "Sound initiation and source types in human imitations of sounds", P. Helgason. In proceedings of FONETIK. pp 83–89, Stockholm (Sweden), 09-11 June, 2014.

43. "Sketch a Scratch." S. Delle Monache, D. Rocchesso, and S. Papetti. 8th International Conference on Tangible, Embedded and Embodied Interaction (TEI), Munich (Germany), 16-19 February, 2014.

Networking:

1. Federico Avanzini, Associate Professor at the University of Padova, Italy `http://www.dei.unipd.it/~avanzini/`

2. Isabelle Ballet, Audio Director at Ubisoft, video game publisher, France, `https://www.ubisoft.com/en-US/studio/paris.aspx`.

3. Julien Bayle, sound and visual artist, creating programmed installations and audiovisual live performances, France, `http://julienbayle.net/`.

4. Laure Bosc, research engineer in digital audio services at PSA Peugeot-Citröen, France.

5. Anthony Lewis Brooks, Associate Professor at Aalborg University Esbjerg, Denmark `http://personprofil.aau.dk/103302`

6. Simon Cacheux, freelance sound designer partime working at PSA Peugeot-Citröen, France, `http://simoncacheux.com/`.

7. Andrea Cera, independent Sound Designer, Italy `http://andrea.cera.free.fr/`.

8. Mark Cartwright, PhD Candidate in Electrical Engineering and Computer Science at Northwestern University - Interactive Audio Lab, Evanston, Illinois, `http://music.cs.northwestern.edu/`.

9. Xavier Collet, freelance sound designer, http://xaviercollet.com.
10. Sebastien Denjean, digital audio engineer at PSA Peugeot-Citröen, France.
11. Carlo Drioli, Assistant Professor at Department of Mathematics and Computer Science, Università degli Studi di Udine, Italy http://people.uniud.it/page/carlo.drioli
12. Richard Dubelski, Performer, Singer, Musician, Paris, France.
13. Cumhur Erkut, Associate Professor at Aalborg University Copenhagen, Denmark http://personprofil.aau.dk/130223
14. Andy Farnell, sound designer, researcher in procedural audio at Queen Mary University of London, England, http://obiwannabe.co.uk/.
15. Remi Forsan, digital audio engineer at PSA Peugeot-Citröen, France.
16. Martin Geijer, Director and Founder at Improvisationsteater Svea AB, Stockholm, Sweden, http://improvisationsteater.se.
17. Christian Heinrichs, PhD candidate in Electronic Engineering and Computer Science at Queen Mary University of London, England.
18. Richard Kronland-Martinet, Research Director at CNRS/LMA - Equipe Sons, Marseille, France, http://www.lma.cnrs-mrs.fr/spip/?lang=fr.
19. Norio Kubo, Director and Product Sound Designer at Yokohama Acoustics Institute, Inc., Yokohama, Japan, http://yokohama-onkyo.jp/.
20. Marc Leman, Professor, Ghent University, Belgium http://research.flw.ugent.be/en/marc.leman.
21. Sylvie-Bobette Levesque, Performer, Singer, Paris, France.
22. Tony Lindeberg, professor, KTH, Dept. of Computational Biology, http://www.csc.kth.se/~tony/.
23. Luca Andrea Ludovico, assistant professor, Università degli Studi di Milano, http://www.ludovico.net/.
24. Martin Luccarelli, Assistant Professor at Department of Design and Art, Libera Università di Bolzano, Bolzano, Italy http://www.unibz.it/en/design-art/welcome/default.html.
25. Luigi Maffei, Full Professor at Department of Architecture and Industrial Design, Seconda Università di Napoli, Napoli, Italy, http://www.architettura.unina2.it/docenti.asp?ID=67.
26. Fernando Ocaña, Creative Director at Semcon Hybrid Design Studios, Sweden, http://www.semcon.com/en/Services/Design/.

27. Elif Ozcan, Assistant Professor, TU Delft, The Netherlands, http://www.io.tudelft.nl/over-de-faculteit/persoonlijke-profielen/universitair-docenten/ozcan-vieira-e/.

28. Stefano Papetti, researcher at Zurich University of the Arts, https://www.zhdk.ch/?person/detail&id=181595.

29. Mathieu Pellerin, freelance Sound Designer, France, http://www.mathieupellerin.com/.

30. Vincent Roussarie, Manager - Human Factors Research and Development Division at PSA Peugeot-Citröen, France.

31. Jean-François Sciabica, scientist and engineer in sound perception at PSA Peugeot-Citröen, France.

32. Stefania Serafin, Professor at Aalborg University Copenhagen, Denmark http://personprofil.aau.dk/107881

33. T. Metin Sezgin, Assistant Professor at Department of Computer Engineering, Koc University College of Engineering, Istanbul, Turkey, http://iui.ku.edu.tr/.

34. Allister Sinclair, freelance sound designer, France.

35. Adam Stark, freelance sound designer and software developer, England, http://www.adamstark.co.uk/.

36. Thibaut Zimmermann, sound designer at PSA Peugeot-Citröen, France.

Dissemination and Communication Activities:

1. "Sound Design Rendezvous", final open event of the SkAT-VG project, Ircam, Paris, 19 January, 2017. http://www.ircam.fr/agenda/le-projet-skat-vg/detail/

2. ".Sound: A Material for Design", interview to the IRCAM Sound Design and Perception Team, by Luc Allemand, Paris, 16 December, 2016. https://www.ircam.fr/article/detail/le-son-materiau-de-design/

3. "Recent researches at IRCAM related to the recognition of rhythm, vocal imitations and music structure", seminar by Geoffroy Peeters (IRCAM) at the Johannes Kepler University in Linz, Austria. December 15th, 2016

4. "Recent researches at IRCAM related to the recognition of rhythm, vocal imitations and music structure", seminar by Geoffroy Peeters (IRCAM) at the Pompeu Fabra University in Barcelona, Spain. November 18th, 2016

5. "Keynote Speech" by Patrick Susini (Ircam) at Audiomostly – http://audiomostly.com/past/norrkoping-sweden-2016/, october 2016, Norrköping, Sweden.

6. "Disegnare il suono con un microfono", IUAV researchers at Venetonight , for the European Researchers' Night in Venice. September 30, 2016.

7. Frédéric Bevilacqua (IRCAM) and Ludovic Germain (sound designer, friend of SkAT-VG) talk about sound design at France Inter radio. September 18, 2016.

8. Workshop on Vocal Sketching by Stefano Delle Monache at the at Acusmatiq festival in Ancona, Italy. July 29, 2016.

9. Lecture by D. Rocchesso at the Virtual Prototyping Summer School. Politecnico di Milano. July 13, 2016.

10. "Sounding Objects and Sonic Sketches", Davide Rocchesso, seminar at the Department of Mathematics and Computer Science, University of Palermo, Italy, http://math.unipa.it/fici/WARG/locandine/2016-07-06_Rocchesso.pdf. July 6, 2016.

11. "48h of sound design at Château La Coste", a short documentary video by Sylvestre Miget, June 2016. `https://vimeo.com/169521601`

12. 48 hours of sound design, Château La Coste. A handful of sound designers sketch the sounds of selected artworks. Final public exhibition. April 28-30, 2016.

13. World Voice Day. 15 April, 2016. Open SkAT-VG Lab at IUAV in Venice.

14. "The Voice as a Professional Tool", symposium on the occasion of the "World Voice Day": six 20-minute lectures and panel in front of a theatre audience at the Stockholm University College of Opera. This included the SkAT-VG presentation by Sten Ternström "Can you imitate something that you have never heard?". Presentations were televised and broadcast on Swedish national educational television several times. Sten Ternström's presentation can be seen at http://urplay.se/program/195860-ur-samtiden-varldsrostdagen-2016-harma-ljud.

15. "Imagining, sketching and prototyping sound", invited talk by D. Rocchesso at the European Centre for Living Technology, Venice, Italy, 4 March, 2016, http://www.unive.it/nqcontent.cfm?a_id=200012.

16. SkAT-VG in FET Through the keyhole, the FET Newsletter of the European Commission. February 2016.

17. "What sound is this gesture?". Guillaume Lemaitre (IRCAM) interviewed on Inside Science. February 2016.

18. "The Skat-VG project : a move to a new sound design tool", open seminar, Ircam, Paris, 27 January, 2016. Recorded sessions – http://medias.ircam.fr/x5a628a.

19. The authors of "Combining gestures and vocalizations during sound imitation", being presented at the 170th ASA Meeting in Jacksonville (USA) have been invited to submit a lay-language version of the paper, which has been posted on the online press room of the Acoustical Society of America, `http://acoustics.org/world-wide-press-room/`, november 2015. An article was published on Gizmodo.

20. "Imagining, sketching and prototyping sound", invited talk by D. Rocchesso at the Sound and Music Computing Colloquium, Aalborg University Copenhagen, 26 November, 2015.

21. Workshop at Aalborg University Copenhagen, Master in Sound and Music Computing, with researchers from IUAV and KTH. 23-27 November 2015.

22. SkAT-VG at ICT2015, Lisbon, Innovate Area, Booth number: i09, 20-22 October, 2015, `http://ec.europa.eu/digital-agenda/events/cf/ict2015/item-display.cfm?id=14872`.

23. "S'i' Fosse Suono", installation by Andrea Cera, October 2015. `http://www.skatvg.eu/SIFosse/`

24. Workshop on Vocal Sketching at York University by IUAV researchers. 07-08 May 2015.

25. Lecture on Sonic Interaction Design by D. Rocchesso at SUPSI (Lugano, CH). 11 December 2014.

26. "Sketching Audio Technologies using Vocalization and Gestures: Le projet SkAT-VG, Projet Européen FP7-ICT FET-Open: challenging current thinking". Séminaire Recherche et Technologie, Ircam, Paris, France, December 2014.

27. Ca' Foscari University Opening of the Doctoral Year. SkAT-VG Exhibit curated by Alan Del Piccolo. 14 November, 2014.

28. "Sketching Audio Technologies using Vocalizations and Gestures". Workshop and demo. Ircam open days, 11 June, 2014, Paris.

29. Research seminar at KTH, the Marcus Wallenberg Laboratories for Sound and Vibration, Pétur Helgason and Sten Ternström presented SkAT-VG and EUNISON projects, 4 April, 2014.

30. World Voice Day. 16 April, 2014. Open seminar "Music and Speech Sounds" at KTH, Stockholm, with several high-profile composers and litterati. SkAT-VG was presented by Sten Ternström and Pétur Helgason.

31. World Voice Day. 16 April, 2014. Performance "Textures from an Exhibition" and SkAT-VG presentation. At IUAV in Venice.

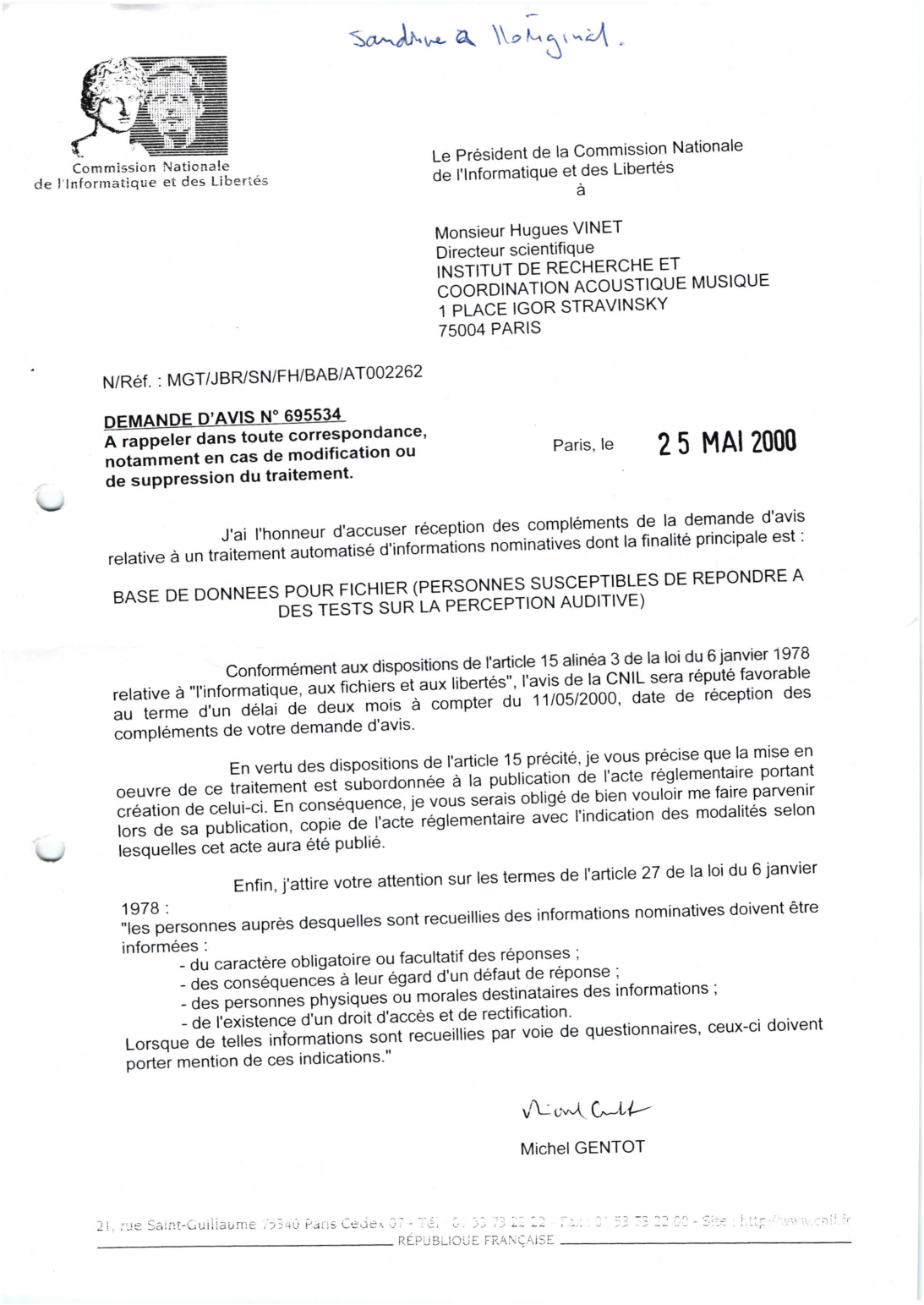
32. Workshop was organized by IUAV at Conservatorio statale di Musica "C. Pollini" in Padova, Italy. 24 March, 2014.

6 Ethical issues

SkAT-VG has been performing psychological experiments, which have been taking place in Paris, within the scope of WP4. A list of subjects for experiments was composed of adult healthy volunteers taken from a database that respects French legislation, and already registered at the CNIL (Commision Nationale de l'Informatique et des Libertés, see Figure 6). This database contains personal information about people (address, age, sex, musical practice) who register themselves to participate in psychological experiments. Subjects had the possibility to leave the experiment any time they wanted. Subjects were reimbursed for their participation, even if they decided to leave the experiment before the end. A new registration for the database of video recordings has been accepted by the CNIL in 2015 (see Figure 7). The experiments performed at IRCAM have been approved by the ethical committee of the French National Institute for Medical Research (CEEI IRB of Inserm) under the number IRB00003888 (pending minor modifications). This committee is a registered Institutional Review Board (IRB) that meets the international ethic standards. A practical consequence is that the results of such experiments can be published in journals such as Plos One, for which an official IRB number is mandatory.

This database is used only within the framework of auditory experiments at IRCAM and will never be shared. The consent form is reproduced in Figure 8. Each subject is identified by a code, and each subject's data is only labeled with this code. The correspondence between subject's code and identity is stored in a separate place.

Within the scope of WP2 and WP3, SkAT-VG has also performed observational studies of performers/imitators. The studies included audio, video and EGG recordings. These experiments have been carried out at KTH, and the participants are improvisational actors recruited through an agency and paid for their participation. Normally, when performing experiments with lay subjects KTH applies for ethical approval from Regionala Etikprövningsnämnden i Stockholm. In this case, these professional stakeholders were requested to sign the consent form reproduced in Figure 9. Ethical approval for making recordings of lay subjects was sought from the regional ethical vetting committee in Stockholm (EPN). In its reply of 2015-09-16, and reported in figure 10, the committee stated that the proposed procedure does not present any ethical objections, and therefore does not need to be considered for approval.



Sandrine à Malgrand.

Commission Nationale
de l'Informatique et des Libertés

Le Président de la Commission Nationale
de l'Informatique et des Libertés
à

Monsieur Hugues VINET
Directeur scientifique
INSTITUT DE RECHERCHE ET
COORDINATION ACOUSTIQUE MUSIQUE
1 PLACE IGOR STRAVINSKY
75004 PARIS

N/Réf. : MGT/JBR/SN/FH/BAB/AT002262

DEMANDE D'AVIS N° 695534
A rappeler dans toute correspondance, notamment en cas de modification ou de suppression du traitement.

Paris, le 25 MAI 2000

J'ai l'honneur d'accuser réception des compléments de la demande d'avis relative à un traitement automatisé d'informations nominatives dont la finalité principale est :

BASE DE DONNEES POUR FICHIER (PERSONNES SUSCEPTIBLES DE REPONDRE A DES TESTS SUR LA PERCEPTION AUDITIVE)

Conformément aux dispositions de l'article 15 alinéa 3 de la loi du 6 janvier 1978 relative à "l'informatique, aux fichiers et aux libertés", l'avis de la CNIL sera réputé favorable au terme d'un délai de deux mois à compter du 11/05/2000, date de réception des compléments de votre demande d'avis.

En vertu des dispositions de l'article 15 précité, je vous précise que la mise en oeuvre de ce traitement est subordonnée à la publication de l'acte réglementaire portant création de celui-ci. En conséquence, je vous serais obligé de bien vouloir me faire parvenir lors de sa publication, copie de l'acte réglementaire avec l'indication des modalités selon lesquelles cet acte aura été publié.

Enfin, j'attire votre attention sur les termes de l'article 27 de la loi du 6 janvier 1978 :
"les personnes auprès desquelles sont recueillies des informations nominatives doivent être informées :
- du caractère obligatoire ou facultatif des réponses ;
- des conséquences à leur égard d'un défaut de réponse ;
- des personnes physiques ou morales destinataires des informations ;
- de l'existence d'un droit d'accès et de rectification.

Lorsque de telles informations sont recueillies par voie de questionnaires, ceux-ci doivent porter mention de ces indications."

Michel GENTOT

21, rue Saint-Guillaume 75340 Paris Cedex 07 - Tél. : 01 53 73 22 22 - Fax : 01 53 73 22 00 - Site : http://www.cnil.fr
RÉPUBLIQUE FRANÇAISE

Figure 6: Acknowledgment of receipt of the IRCAM database by the Commission Nationale de l'Informatique et des Libertés (CNIL). According to this letter, no answer within two months indicates that the database is accepted.

CNIL
8 rue de Vivienne - 75083 PARIS cedex 02
T. 01 53 73 22 22 - F. 01 53 73 22 00
www.cnil.fr

RÉCÉPISSÉ

DÉCLARATION NORMALE

Numéro de déclaration
1890949 v 0
du 25 septembre 2015

Monsieur LEMAITRE Guillaume
INSTITUT DE RECHERCHE ET DE
COORDINATION ACOUSTIQUE MUSIQUE
PERCEPTION ET DESIGN SONORES
1 PLACE STRAVINSKY
75004 PARIS

A LIRE IMPERATIVEMENT
La délivrance de ce récépissé atteste que vous avez transmis à la CNIL un dossier de déclaration formellement complet. Vous pouvez désormais mettre en oeuvre votre traitement de données à caractère personnel.
La CNIL peut à tout moment vérifier, par courrier, par la voie d'un contrôle sur place ou en ligne, que ce traitement respecte l'ensemble des dispositions de la loi du 6 janvier 1978 modifiée en 2004. Afin d'être conforme à la loi, vous êtes tenu de respecter tout au long de votre traitement les obligations prévues et notamment :
1) La définition et le respect de la finalité du traitement,
2) La pertinence des données traitées,
3) La conservation pendant une durée limitée des données,
4) La sécurité et la confidentialité des données,
5) Le respect des droits des intéressés : information sur leur droit d'accès, de rectification et d'opposition.
Pour plus de détails sur les obligations prévues par la loi « informatique et libertés », consultez le site internet de la CNIL : www.cnil.fr

Organisme déclarant

Nom : INSTITUT DE RECHERCHE ET DE COORDINATION ACOUSTIQUE MUSIQUE
Service : DEPARTEMENT R&D
Adresse : 1 PLACE STRAVINSKY
Code postal : 75004
Ville : PARIS

N° SIREN ou SIRET : 309320612 00018
Code NAF ou APE : 7112B
Tél. : 01 44 78 13 50
Fax. :

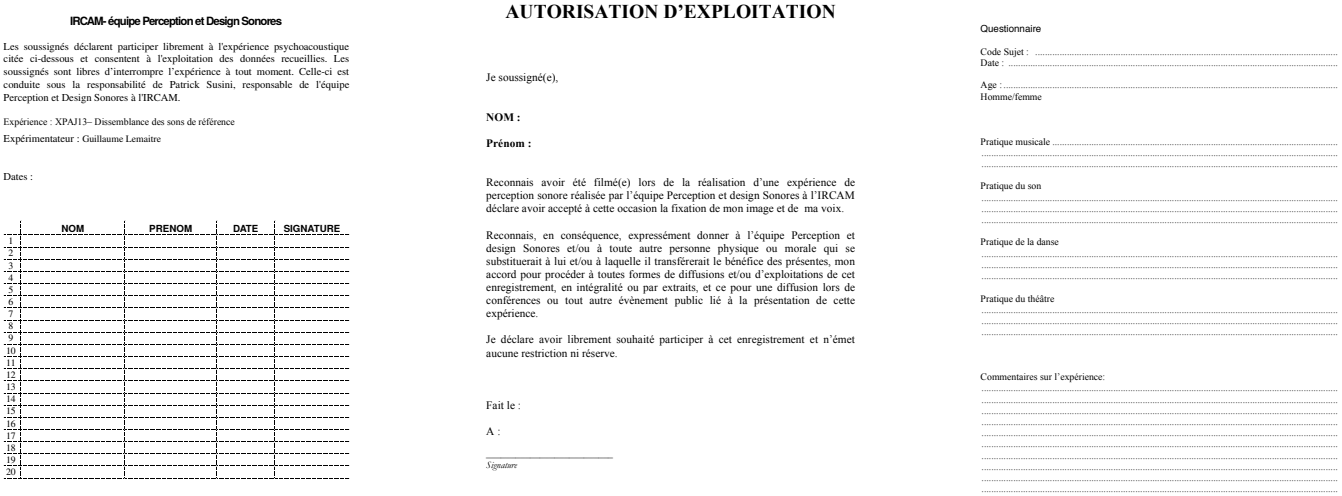
Traitement déclaré

Finalité : NOUS EFFECTUONS DES RECHERCHES SCIENTIFIQUES SUR LA PERCEPTION DES SONS, LA VOIX, ET LES GESTES. NOUS EFFECTUONS DES ENREGISTREMENTS AUDIO ET VIDEOS DE PERSONNES DECRIVANT OU IMITANT DES SONS (AVEC LEUR CONSENTEMENT). LES DONNEES INFORMATIQUES SONT ANONYMES, MAIS NOUS CONSERVONS, DE MANIERE SEPAREE EN FORMAT PAPIER, LE CONSENTEMENT DES PARTICIPANTS, ET DONC LES COORDONNEES ET L'ETAT CIVIL DE CES PARTICIPANTS.

Fait à Paris, le 25 septembre 2015
Par délégation de la commission

Isabelle FALQUE PIERROTIN
Présidente

Figure 7: Acknowledgment of receipt of the IRCAM database of video recordings (2015) by the Commission Nationale de l'Informatique et des Libertés (CNIL)



IRCAM- équipe Perception et Design Sonores

Les soussignés déclarent participer librement à l'expérience psychoacoustique citée ci-dessous et consentent à l'exploitation des données recueillies. Les soussignés sont libres d'interrompre l'expérience à tout moment. Celle-ci est conduite sous la responsabilité de Patrick Susini, responsable de l'équipe Perception et Design Sonores à l'IRCAM.

Expérience : XPAJ13– Dissemblance des sons de référence

Expérimentateur : Guillaume Lemaitre

Dates :

	NOM	PRENOM	DATE	SIGNATURE
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

AUTORISATION D'EXPLOITATION

Je soussigné(e),

NOM :

Prénom :

Reconnais avoir été filmé(e) lors de la réalisation d'une expérience de perception sonore réalisée par l'équipe Perception et design Sonores à l'IRCAM déclare avoir accepté à cette occasion la fixation de mon image et de ma voix.

Reconnais, en conséquence, expressément donner à l'équipe Perception et design Sonores et/ou à toute autre personne physique ou morale qui se substituerait à lui et/ou à laquelle il transférerait le bénéfice des présentes, mon accord pour procéder à toutes formes de diffusions et/ou d'exploitations de cet enregistrement, en intégralité ou par extraits, et ce pour une diffusion lors de conférences ou tout autre évènement public lié à la présentation de cette expérience.

Je déclare avoir librement souhaité participer à cet enregistrement et n'émet aucune restriction ni réserve.

Fait le :

A :

Signature

Questionnaire

Code Sujet : .....
Date : .....

Age : .....
Homme/femme

Pratique musicale .....

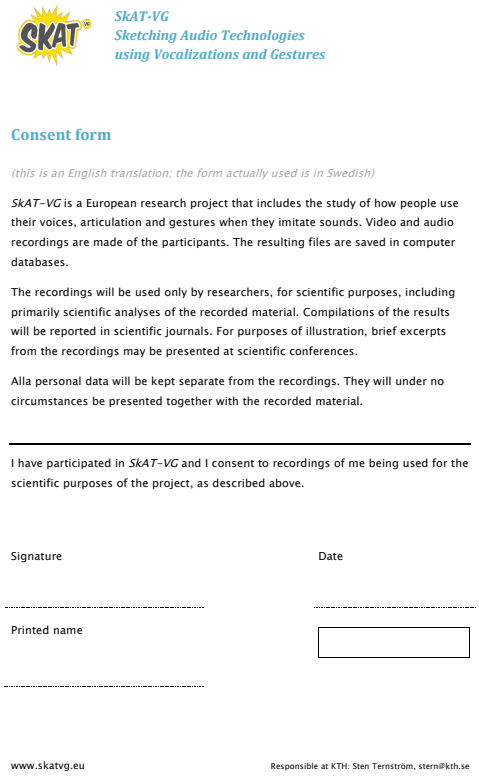
Pratique du son

Pratique de la danse

Pratique du théâtre

Commentaires sur l'expérience:

Figure 8: IRCAM Consent forms and questionnaire for subjects.



SkAT-VG
Sketching Audio Technologies
using Vocalizations and Gestures

Consent form

(this is an English translation; the form actually used is in Swedish)

SkAT-VG is a European research project that includes the study of how people use their voices, articulation and gestures when they imitate sounds. Video and audio recordings are made of the participants. The resulting files are saved in computer databases.

The recordings will be used only by researchers, for scientific purposes, including primarily scientific analyses of the recorded material. Compilations of the results will be reported in scientific journals. For purposes of illustration, brief excerpts from the recordings may be presented at scientific conferences.

Alla personal data will be kept separate from the recordings. They will under no circumstances be presented together with the recorded material.

I have participated in *SkAT-VG* and I consent to recordings of me being used for the scientific purposes of the project, as described above.

Signature

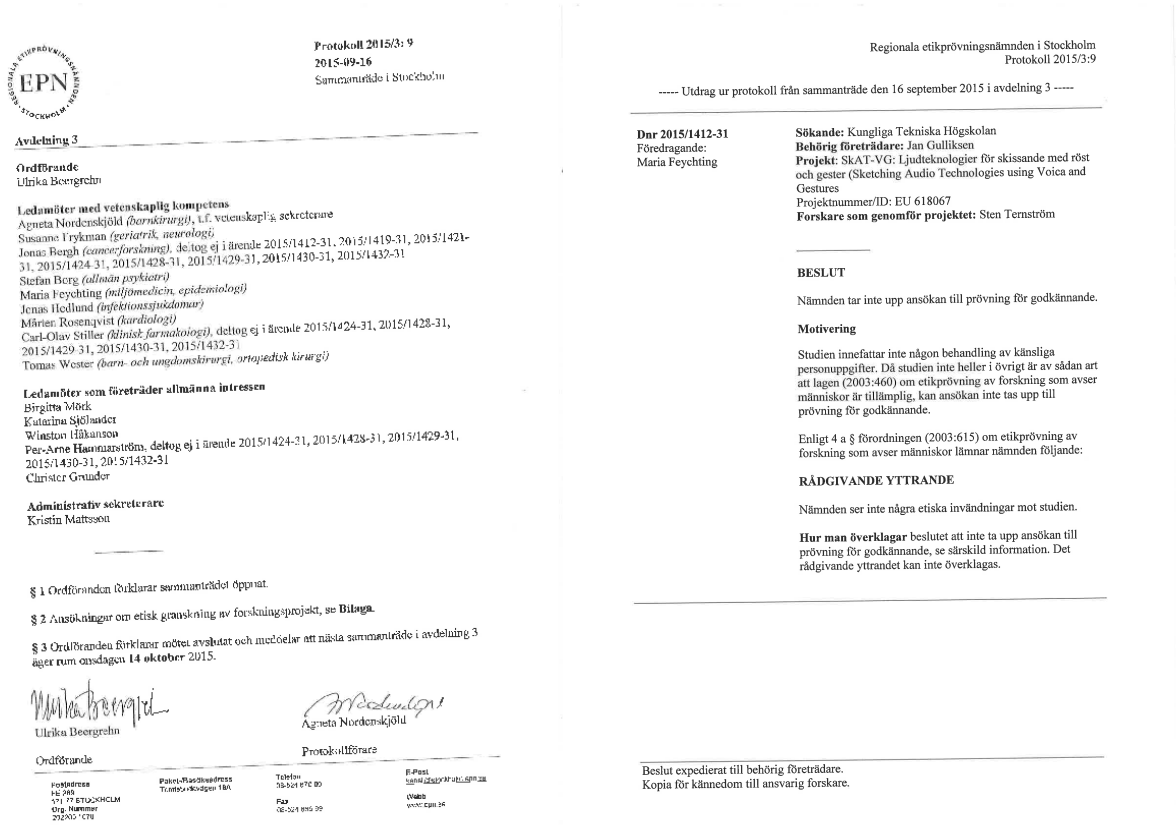
Date

Printed name

www.skatvg.eu

Responsible at KTH: Sten Ternström, stern@kth.se

Figure 9: KTH consent form.



EPN

Protokoll 2015/3: 9
2015-09-16
Sammanträde i Stockholm

Avdelning 3

Ordförande
Ulrika Beergrehn

Ledamöter med vetenskaplig kompetens
Agneta Nordenskjöld *(barnkirurgi)*, t.f. vetenskaplig sekreterare
Susanne Frykman *(geriatrik, neurologi)*
Jonas Bergh *(cancerforskning)*, deltog ej i ärende 2015/1412-31, 2015/1419-31, 2015/1421-31, 2015/1424-31, 2015/1428-31, 2015/1429-31, 2015/1430-31, 2015/1432-31
Stefan Berg *(allmän psykiatri)*
Maria Feychting *(miljömedicin, epidemiologi)*
Jonas Hedlund *(infektionssjuksköterska)*
Mårten Rosenqvist *(kardiologi)*
Carl-Olav Stiller *(klinisk farmakologi)*, deltog ej i ärende 2015/1424-31, 2015/1428-31, 2015/1429-31, 2015/1430-31, 2015/1432-31
Tomas Wester *(barn- och ungdomskirurgi, ortopedisk kirurgi)*

Ledamöter som företräder allmänna intressen
Birgitta Mörk
Katarina Skånhader
Winston Håkanson
Per-Arne Hammarström, **deltog ej** i ärende 2015/1424-31, 2015/1428-31, 2015/1429-31, 2015/1430-31, 2015/1432-31
Christer Grundér

Administrativ sekreterare
Kristin Mattsson

§ 1 Ordföranden förklarar sammanträdet öppnat.

§ 2 Ansökningar om etisk granskning av forskningsprojekt, se **Bilaga**.

§ 3 Ordföranden förklarar mötet avslutat och meddelar att nästa sammanträde i avdelning 3 äger rum onsdagen **14 oktober** 2015.

Ulrika Beergrehn
Ordförande

Agneta Nordenskjöld
Protokollförare

Postadress
Box 2148
171 77 STOCKHOLM
Org. Nummer
202100-7778

Paket-/Besöksadress
Tomtebodavägen 18A

Telefon
08-524 878 80

Fax
08-524 878 99

E-Post
registrator@epn.se

Webb
www.epn.se

Regionala etikprövningsnämnden i Stockholm
Protokoll 2015/3:9

----- Utdrag ur protokoll från sammanträde den 16 september 2015 i avdelning 3 -----

Dnr 2015/1412-31
Föredragande:
Maria Feychting

Sökande: Kungliga Tekniska Högskolan
Behörig företrädare: Jan Gulliksen
Projekt: SkAT-VG: Ljudteknologier för skissande med röst och gester (Sketching Audio Technologies using Voica and Gestures
Projektnummer/ID: EU 618067
Forskare som genomför projektet: Sten Ternström

BESLUT

Nämnden tar inte upp ansökan till prövning för godkännande.

Motivering

Studien innefattar inte någon behandling av känsliga personuppgifter. Då studien inte heller i övrigt är av sådan art att lagen (2003:460) om etikprövning av forskning som avser människor är tillämplig, kan ansökan inte tas upp till prövning för godkännande.

Enligt 4 a § förordningen (2003:615) om etikprövning av forskning som avser människor lämnar nämnden följande:

RÅDGIVANDE YTTRANDE

Nämnden ser inte några etiska invändningar mot studien.

Hur man överklagar beslutet att inte ta upp ansökan till prövning för godkännande, se särskild information. Det rådgivande yttrandet kan inte överklagas.

Beslut expedierat till behörig företrädare.
Kopia för kännedom till ansvarig forskare.

Figure 10: Reply from EPN-Stockholm

7 Relations with other projects

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 wavelet. The analysis of the different multimodal strategies the listeners use to describe sound was performed entirely within SkAT-VG. The classification work of WP5 has been largely based on the findings of WP3 and WP4, well within the boundary of SkAT-VG, and brand new features have been derived for the project purposes. The classification task in SkAT-VG has led to the extension of an existing framework. Thanks to SkAT-VG funds the automatic classifiers now model the evolution in time of sounds, thus tackling a different, and 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. 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 WP3 and WP4, 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 sound designers interviews. 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 8 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 8, have been entirely developed or collected in SkAT-VG.

As a side note, a detailed list of Background Included is provided in the Consortium Agreement, to provide access rights to Background made available to the Parties. It also provides factual information about the state of the art upon which the Consortium has been building its research.

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 8: Relations with other projects.

References

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

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

[BLDB15] Stefano Baldan, Hélène Lachambre, Stefano Delle Monache, and Patrick Boussard. Physically informed car engine sound synthesis for virtual and augmented environments. In *Proceedings of the 2nd Workshop on Sonic Interactions in Virtual Environments - IEEE VR 2015*, Arles, France, 2015.

[CMR16] Andrea Cera, Davide Andrea Mauro, and Davide Rocchesso. Sonic in(tro)spection by vocal sketching. In *Proceedings of the XXI CIM*, Cagliari, Italy, september 2016. AIMI.

[DBM+17] Clemént Dendievel, Stefano Baldan, Stefano Delle Monache, Hélène Lachambre, Davide Rocchesso, and Patrick Boussard. Understanding and modifying sound design practices. 2017. In preparation.

[DBMR14] Stefano Delle Monache, Stefano Baldan, Davide Andrea Mauro, and Davide Rocchesso. A design exploration on the effectiveness of vocal imitations. In *Proc. of the Sound and Music Computing Conference*, Athens, Greece, September 2014.

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

[DR16a] Stefano Delle Monache and Davide Rocchesso. Cooperative sound design: A protocol analysis. In *Proc. of the Audio Mostly 2016*, AM '16, pages 154–161, New York, NY, USA, 2016. ACM.

[DR16b] Stefano Delle Monache and Davide Rocchesso. Understanding cooperative sound design through linkographic analysis. In *Proceedings of the XXI CIM Colloquium on Music Informatics*, Cagliari, Italy, 2016.

[DRBM15] Stefano Delle Monache, Davide Rocchesso, Stefano Baldan, and Davide Andrea Mauro. Growing the practice of vocal sketching. In *Proceedings of the 21st International Conference on Auditory Display (ICAD 2015)*, pages 58 – 65, Graz, Austria, 2015.

[DRP14] Stefano Delle Monache, Davide Rocchesso, and Stefano Papetti. Sketch a scratch. In *Eight International Conference on Tangible, Embedded and Embodied Interaction*, TEI '14, 2014.

[ERDS16] Cumhur Erkut, Davide Rocchesso, Stefano Delle Monache, and Stefania Serafin. A case of cooperative sound design. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, NordiCHI '16, pages 83:1–83:6, New York, NY, USA, 2016. ACM.

[Hel14] Pétur Helgason. Sound initiation and source types in human imitations of sounds. In *Proceedings FONETIK*, pages pp 83–89, Stockholm (Sweden), 09-11 June 2014.

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

[ISL16] Vincent Isnard, Clara Suied, and Guillaume Lemaitre. Auditory bubbles reveal sparse time-frequency cues subserving identification of musical voices and instruments. *The Journal of the Acoustical Society of America*, 140(4):3267, 2016. Proceedings of the meeting of Acoustical Society of America, Honolulu, HI.

[LF15a] Tony Lindeberg and Anders Friberg. Idealized computational models for auditory receptive fields. *PloS one*, 10(3), 2015.

[LF15b] Tony Lindeberg and Anders Friberg. Scale-space theory for auditory signals. In *Scale Space and Variational Methods in Computer Vision*, pages 3–15. Springer, 2015.

[LHV+16] Guillaume Lemaitre, Olivier Houix, Frédéric Voisin, Nicolas Misdariis, and Patrick Susini. Comparing the identification of vocal imitations and computational sketches of everyday sounds. *The Journal of the Acoustical Society of America*, 140(4):3390, 2016. Proceedings of the meeting of Acoustical Society of America, Honolulu, HI.

[LHV+17] Guillaume Lemaitre, Olivier Houix, Frédéric Voisin, Nicolas Misdariis, and Patrick Susini. Vocal imitations of non-vocal sounds. 2017. Plos One (in press).

[LJH+15] Guillaume Lemaitre, Ali Jabbari, Olivier Houix, Nicolas Misdariis, and Patrick Susini. Vocal imitations of basic auditory features. *The Journal of the Acoustical Society of America*, 137(4):2268, 2015. Proceedings of the meeting of Acoustical Society of America, Pittsburgh, PA.

[LJM+16] Guillaume Lemaitre, Ali Jabbari, Nicolas Misdariis, Olivier Houix, and Patrick Susini. Vocal imitations of basic auditory features. *The Journal of the Acoustical Society of America*, 139(1):290–300, 2016.

[LR14] Guillaume Lemaitre and Davide Rocchesso. On the effectiveness of vocal imitations and verbal descriptions of sounds. *The Journal of the Acoustical Society of America*, 135(2):862–873, 2014.

[LSFB] Guillaume Lemaitre, Hugo Scurto, Jules Françoise, and Frédéric Bevilacqua. Imitating sounds with gestures. Manuscript in preparation.

[MP15] Enrico Marchetto and Geoffroy Peeters. A set of audio features for the morphological description of vocal imitations. In *Proceedings of the International Conference on Digital Audio Effects*, Trondheim, Norway, 2015.

[New04] Fred Newman. *MouthSounds: How to Whistle, Pop, Boing, and Honk... for All Occasions and Then Some*. Workman Publishing, 2004.

[PBO+14] G. Palmas, M. Bachynskyi, A. Oulasvirta, H. P. Seidel, and T. Weinkauf. An edge-bundling layout for interactive parallel coordinates. In *2014 IEEE Pacific Visualization Symposium*, pages 57–64, March 2014.

[PR16] Alan Del Piccolo and Davide Rocchesso. Non-speech voice for sonic interaction: a catalogue. *Journal on Multimodal User Interfaces*, pages 1–17, 2016.

[RBB+17] Davide Rocchesso, Stefano Baldan, Frederic Bevilacqua, Andrea Cera, Stefano Delle Monache, and Guillaume Lemaitre. Embodied sound design. 2017. Submitted.

[RDA16] Davide Rocchesso, Stefano Delle Monache, and Mauro Davide A. miMic: The microphone as a pencil. In *Tenth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '16, 2016.

[RMP16] D. Rocchesso, S. Delle Monache, and S. Papetti. Multisensory texture exploration at the tip of the pen. *International Journal of Human-Computer Studies*, 85:47 – 56, 2016.

[SLF+15] Hugo Scurto, Guillaume Lemaitre, Jules Françoise, Frédéric Voisin, Frédéric Bevilacqua, and Patrick Susini. Combining gestures and vocalizations to imitate sounds. *The Journal of the Acoustical Society of America*, 138(3):1780–1780, 2015. Proceedings of the meeting of the Acoustical Society of America, Jacksonville, FL.

[SLF+16] Hugo Scurto, Guillaume Lemaitre, Jules Françoise, Frédéric Bevilacqua, Patrick Susini, and Frédéric Voisin. Embodying sounds: Building and analysis of a database of gestural and vocal imitations. In *Gesture, Creativity, Multimodality. Procedings of ISGS 7 : 7^{th} Conference of the International Society for Gesture Studies, Paris.*, page 269, The University of Texas at Austin, Austin, TX, 2016. International Society for Gesture Studies.