miMic: The Microphone as a Pencil

Davide Rocchesso, Davide A. Mauro, and Stefano Delle Monache
Iuav University of Venice, Department of Architecture and Arts
Dorsoduro 2206, Venice, Italy
{roc, dmauro, sdellemonache}@iuav.it

ABSTRACT
miMic, a sonic analogue of paper and pencil is proposed: An augmented microphone for vocal and gestural sonic sketching. Vocalizations are classified and interpreted as instances of sound models, which the user can play with by vocal and gestural control. The physical device is based on a modified microphone, with embedded inertial sensors and buttons. Sound models can be selected by vocal imitations that are automatically classified, and each model is mapped to vocal and gestural features for real-time control. With miMic, the sound designer can explore a vast sonic space and quickly produce expressive sonic sketches, which may be turned into sound prototypes by further adjustment of model parameters.

ACM Classification Keywords
H.5.5 Sound and Music Computing: Systems

Author Keywords
Vocal Sketching; Gestures; Sonic Interaction Design; Augmented Microphone

INTRODUCTION
Architects, product designers, graphic designers, and several other kinds of professionals embracing design with a visual attitude, approach the early stages of the design process armed with paper and pencil [12], or sometimes with electronic equivalents when they are available [3, 13]. This is not the case when it comes to designing the non-visual aspects of products, especially sound, where there is no obvious analogue of paper and pencil. It is true that music composers often call sketches what are fragments of written music that may lead to or be used in a composition. That is indeed very similar to sketching in visual design, but has little to do with the perceptual qualities of sounds, which are appreciated through the ears. One may think that the audio recorder may be the most direct audio sketching tool. The recorder allows one to store sonic ideas in audible form, but it lacks the flexibility, expressiveness, and immediacy that drawing on paper brings about. In fact, sketches evolve by addition, deletion, or overlays, until they are ready to be 'hard-lined' [12].

Figure 1. miMic in use.

If drawing exploits the motor abilities of human hands, what kind of human abilities should a sound sketching tool exploit, in order to promote the qualities of sketching in sound design? And, before that, what are the qualities of sketching in the auditory domain?

Following Buxton [3], we assume that sketches are quick, timely, inexpensive, disposable, and plentiful. They have a clear, gestural vocabulary, and are proposed with constrained resolution and appropriate degree of refinement. Sketches are intentionally ambiguous and open to interpretation, and they suggest exploration. All of these qualities apply to sketching in any domain, but in the case of sound they are naturally associated with what humans have always been doing to communicate sonic phenomena, i.e., using voice and gesture. Children sketch sound before acquiring language, and adults easily resort to vocal sketches when they have no words to communicate a sonic concept [18]. So, if voice and hand are our natural audio-visual sketching apparati, how do we provide paper and pencil to sketch sounds?

TEI '16, February 14-17, 2016, Eindhoven, Netherlands
ACM 978-1-4503-3582-9/16/02.
http://dx.doi.org/10.1145/2839462.2839467

We propose that the microphone can become the pencil for sound, especially if it is augmented to capture gestures, and if it affords direct access and action on a wide palette of sound models. We present and demonstrate miMic, the sketching microphone (Figure 1).

The analogy between vocal sketching and drawing was proposed in the context of sonic interaction design [9]. Vocal sketches are the most direct representations of concepts in the aural domain. They are immediately available and do not suffer from any technological constraint, if not of the natural conformation of the vocal apparatus. On the other side, one major design drawback of drafting sound ideas by using the voice is the inherent ephemerality of this display, especially when the designer needs to move on in the working process, towards the refinement and prototyping stages. So far, no effective tools have been developed, that exploit the power of the human voice in the early stages of the sound design process, whereas the effectiveness of communication of audio concepts via vocal imitations has been experimentally assessed, especially for sound retrieval [7, 2]. Conversely, several examples can be found in the literature of systems that implement the vocal sampler-looper-effector chain for prototyping or for producing music patterns [29]. More relevant and inspiring for the reported research are the studies that attempted at extracting acoustic or phonetic qualities from the voice to drive sound synthesis for musical purposes [16, 27].

We are not the first to propose augmented microphones, although the examples found in the literature were mainly devoted to extended vocal performance. In recent years, a prototype microphone with sensors and switches was proposed by Sennheiser [24], and a microphone for voice augmentation and modification was presented by Park et al. [23]. Much earlier, a sensorized microphone stand was proposed to capture the singer performance gestures [14]. Other projects[1] tried to tie voice recording and gesture, so that the recorded vocalizations can be gesturally manipulated.

In this article we are proposing to embed an Inertial Measurement Unit (IMU) and two switches into a conventional microphone. The goal is to introduce a system architecture that, through the augmented microphone, can empower the sound designer with a wide sound palette that can be directly controlled by voice and gesture. As seen in a long term vision, the proposed system architecture is being developed with the aim of providing effective tools for sound designers, to support the conceptual stage of the design process.

We first describe how the design of miMic matured through workshops, discussions, and shared documentation. Then, we present the realization of the physical part of miMic. Next we discuss the modes of use of the tool and we briefly describe the synthesis and control part of the system architecture.

DESIGN

The design rationale for miMic emerged out of several experiences in the community of sonic interaction design. To our knowledge, the first workshop on vocal sketching for sonic interactions was run in Israel in 2009 [9]. After that, there has been a growing corpus of experiences derived from workshops combining vocal sketching, Foley artistry, interactive-object manipulation, and sound synthesis [5, 20].

Workshops on Vocal Sketching

Despite the scarce literature available on vocal sketching, this approach is becoming progressively popular in the educational and research domain, as alternative means to explore and communicate sound design concepts. Introductory and warm-up exercises are available, to foster the practice and training of vocal imitations. Typically, a playful and collaborative playground is set in order to encourage improvisation, engagement, and immediacy in the production of voice-driven displays. The assignments may range from the free exploration of vocal techniques [21], and the representation of sonic memories [5, 20], to the proper design of displays and product sounds [9, 8].

Based on these aforementioned experiences, we devised the system architecture of miMic to possibly overcome the critical points [15], and to provide a computational tool that would seamlessly support the ubiquity of sketching.

Body-Storming

The devised set of features of miMic system architecture emerged during a body-storming session with five participants, including the authors. We exploited concurrent and retrospective think-aloud protocols in the execution of simple design tasks, to highlight interaction styles with the microphone and modes of use, in terms of expectations and control opportunities for the purpose of sonic sketching. In practice, one performer played the role of the miMic's system, responding with his own voice to the vocalizations and gestures of the sound designer, hence interpreting the interaction with the augmented microphone.

Two main aspects emerged from the session analysis, one related to the emerging sequences of the sound design process, and one related to the desired interactive behavior of miMic. We observed a process that could be roughly split into four stages: 0 Communication, by vocal examples and verbal specification, of a sound type; 1. Personification of the sound type, with the designer mimicking a sound behavior and the system-person mirroring it; 2. Expansion of the sonic space by exploration of new vocal control territories; 3. Saturation of the non-verbal conversation between designer and system, when additional information need to be exchanged to achieve proper refinement of the sonic sketch. Stage 2 was actually made possible by the distinction between stages 0 and 1, as the selection of a certain sound family, represented by a sound model, allows the designer to exploit the whole range of vocal possibilities to control the given model. For example, someone produced bubble-like sounds to control a combustion-engine sound model, thus clearly exceeding the boundary of imitative control of the sound model.

A further important observation concerns the emerging use of manual gestures in the interaction with the augmented microphone. Indeed, the original hypothesis was to capture the gestures by means of a micro-camera embedded in the microphone shell. However, during the body-storming session,

[1] e.g., VOGST: http://blogs.iad.zhdk.ch/vogst/

this solution produced the effect of decoupling voice from gestures. For this reason, we opted for focusing on those gestures that could be achieved by handling the microphone with one hand, as depicted in Figure 1. In this way, we could exploit the nuances of subtle movements, such as shaking and spinning the microphone, to control additional parameters of the sound model. Eventually, this control strategy gives access to the manipulation of sound morphologies that are not immediately or intuitively controllable with the voice.

The Four Stages of Sound Design

The experience gained through workshops, brainstorming session, tests with partial realizations, and further discussions, led us to foresee a sound design process structured into four stages and three modes of use of the proposed system architecture, as described in Table 1.

n.	Stage	Mode	Tool
0	Select	Select	miMic
1	Mimic	Play	miMic
2	Explore		miMic
3	Refine	Play + Tune	miMic + GUI

Table 1. The four stages of sound design, using miMic. The Refine stage implies the Play mode of use of miMic to be complemented with manual operation on the GUI (Tune).

miMic is the exclusive tool in three of the four sound design stages, and in the fourth stage it complements the GUI elements that sound designers normally use. The four stages of sound design can be made to correspond roughly to the four iterations of the 'design funnel' [3, 13]: concepts, exploration, clarification, resolution. In stage 0 of our sound design funnel, by selecting one sound model or a mixture of sound models, the designer is effectively defining a sonic concept [22]. At stage 1, by vocally mimicking a selected sound mechanism, she gets acquainted with the available sound space, simply by vocal mirroring. Exploration is extended at stage 2, as soon as the user realizes that control is not limited to a restricted set of vocalizations and gestures. One can go beyond mirroring and let creative uses of voice and gestures explore new neighborhoods of the sonic space that is made available by a given sound model. While stages 0 to 2 can all be performed through the miMic without any visual display, stage 3 possibly requires the manipulation of each single model parameter, made available as a GUI element (typically, a virtual slider). miMic can still be supportive in this stage, as real-time vocal-gestural control can be applied while changing parameters one by one. Stage 3 is indeed the resolving iteration [3], where a sound sketch is hard-lined into a sound prototype.

A note should be made on stage 0, where a choice is made from a relatively small set of available models. This initial step does not aim at reducing the sound design space, but it rather stems from the observation that sketches are widely used to typify information. This is a 'seeing as' move, as theorized by Goldschmidt [12], and it is followed by 'seeing that' moves, where generic qualities are translated into specific appearances. Moreover, selection can be made to produce a mixture of sound models, weighted according to the likelihoods returned by the classifier. In this way, the boundaries between different models become effectively softer.

The Product as a Platform

The system architecture supporting the intuitive use of miMic as a sketching tool requires the joint development of several components. In this paper, we separately address the physical device in the Tool section, and the software components supporting the modes of use in the Select, Play, and Gesticulate sections, respectively. The hardware-software ensemble is offered as a platform product [4], open to extensions and derivatives of any kind. The whole design process, complete with source files, is being documented in the Build in Progress platform[2] [28].

TOOL

The physical device for vocal sketching is an augmented microphone. Apart from the usual capability of converting sound pressure waves into audio signals, miMic has been conceived to sense the manual gestures exerted on the microphone and convert them into control signals.

Manual Gestures

The embedded IMU captures acceleration and orientation signals that can be used for continuous gesture exploitation. Manual gestures often accompany vocal imitations and are normally used to reinforce and complement vocalizations [26]. Acceleration and orientation signals can be directly coupled to sound model parameters, or gesture dynamic qualities can be extracted from them [1]. Relevant qualities can be the gesture energy, or the degree of smoothness, as well as the timing of major strokes.

Modes of Use

During the conception of miMic, especially during workshops and bodystorming session, the two modes Select and Play emerged, the first for choosing among the palette of sound models, and the second for playing with the selected model(s). In principle, it would be possible to achieve a modeless interface by implementing a seamless model selection system. This would mean that, for example, if the user starts mimicking a car engine the corresponding sound synthesis would be activated and it would be possible to control it with vocal features. However, it was found that this option would limit the creative process of sound design and make stage 2 ("Explore", in Table 1) very difficult to achieve, as the vocal control space would need to be as limited as the sound synthesis space. In fact, the designer would not be allowed to make, for example, turbulent noisy sounds with her voice to control the engine sound, as this would immediately activate a wind-like sound model.

To distinguish between Select and Play, a single button would be sufficient, and to avoid modal errors, a quasi-modal interface could be obtained by using a spring-loaded button. Instead, we opted for two distinct latching buttons with embedded LED for light feedback, so that (i) the current mode is in the user's locus of attention, and (ii) two other states are

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

available for two further operations. In the current realization, only one of these two extra states is used: The model selection process (stage 0, Select) can be personalized with the user-provided vocal imitations, by activating the Train procedure when the two buttons are simultaneously pushed. This is a one-off operation that the sound designer would perform once or rarely.

The two buttons serve as the only visible interface for the user. The reason is to hide, in stages 0 to 2 (Select, Mimic, Explore) of the sound design process, all the complexity of model selection and sound model parameters, thus promoting a creative exploration of the sonic space.

Shape

The main questions in giving form to the physical tool are: What kind of microphone shape? Where to put the buttons?

To give an answer to these questions, we decided that the microphone should be graspable with a single hand (like a pencil is) and actuated with a couple of fingers. Given these requirements the attention focused on stage microphones, thus excluding studio configurations that are not designed to be manipulated. The two possible shapes are: "gelato"-style (Shure-SM58 *et similia*), or "vintage"-style (Shure-55 *et similia*).

For the gelato shape, there is ample possibility to accomodate buttons on the stick, as in the Sennheiser Concept Tahoe [24]. However, the vintage roundish shape is preferred as: It compactly fits one hand, it is an unequivocal visual icon, it can sit vertical on a plane, it can fit two large visible buttons on top, and electronics inside.

Construction and Components

The components used in the project are: Microphone (to be hacked) Soundsation TA-54D[3]; push buttons, latching, with light[4] (one white, one blue); IMU Adafruit LSM9DS0[5]; microcontroller board Arduino Nano[6]; jumpers, wires, and two 220 Ω resistors; segments of metal tube.

The two buttons have been put on top of the frontal shell. Two holes have been drilled and pieces of metal tube have been used to raise the buttons from the shell. An Adafruit tutorial contains all information for wiring the IMU to the microcontroller board[7]. The current prototype is doubly wired, with both an audio cable and a USB cable. To keep the parts easily removable, soldering has been limited to a minimum, and jumper wires have been used.

SELECT

Selection of one or more sound models is operated by automatic classification of vocal imitations. There are two different approaches to the design of this function:

People centered: The tool is supposed to work for the casual user, based on what the classifier learned from many imitations provided by a large pool of subjects.

Individual centered: The classifier is trained to recognize the imitations of a specific user.

People-Centered Selection

To make a sound model selection that works for the layperson, a machine classifier should be trained on a significantly large sample of vocalizations, labeled according to perceptually-meaningful classes. Such collection and labeling of vocal samples is a challenging task per se, which is indeed being accomplished in the project SkAT-VG[8]. In order to reach a useful level of accuracy (for example, beyond 70% in selection among a dozen classes), specific morphological audio descriptors have been developed [19].

In this work we do not focus on the performance of voice-based sound-model selection, but rather on the implications that different kinds of selection have on system design. To demonstrate the people-centered Select mode of miMic we implemented a basic sound model selector based on a classification tree. The construction of such classifier is based on the following steps: 1. Collect examples; 2. Train a classifier (offline); 3. Implement an online recognition system.

For the present realization of miMic, we have been considering the following classes of sounds:

DC – electric motors (265 examples);

Engine – internal combustion motors (257 examples).

Liquid – fluid-dynamic noises (275 examples);

Saw – scraping or sawing (271 examples);

Wind – aerodynamic noises (261 examples).

The collected examples of these classes sum up to a total of 1329 samples, which have been taken from the 8000-imitations database of project SkAT-VG. This amounts to roughly one sixth of the whole database, which is representative of a large repertoire of vocal imitations. The examples have been amplitude normalized at −1dBFS, and their length is at least 4 seconds.

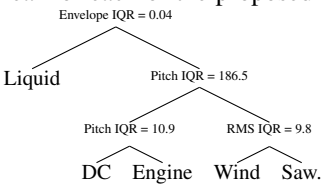
Offline Training

For each sample in the database the following set of features have been extracted under Cycling'74 Max: 1. Centroid; 2. Variance; 3. Skewness; 4. Kurtosis; 5. Flatness; 6. Flux; 7. Onset; 8. Pitch; 9. Envelope; 10. RMS.

For each feature, computed on windows of 4096 samples, with an overlap of 75%, the Median and InterQuartile Range (IQR) were computed over the sample length. The Ratio between IQR and Median was computed for Envelope and RMS features since they are dependent on the signal level. The Cycling'74 Max patch produced a line of text for each imitation example, including the label of the class and the sequence of feature values. A binary classification tree was derived by using the Matlab `fitctree` function,

[3] http://www.soundsationmusic.com/?p=25891
[4] https://www.sparkfun.com/products/11975
[5] http://www.adafruit.com/products/2021
[6] http://arduino.cc/en/Main/arduinoBoardNano
[7] https://learn.adafruit.com/adafruit-lsm9ds0-accelerometer-gyro-magnetometer-9-dof-breakouts/overview
[8] http://www.skatvg.eu

and severely pruned to improve generalization. To obtain a balance between interpretability and precision of this tree we chose the pruning depth that allows to have at least one leaf for each of the proposed models, thus obtaining the tree



Online Model Selection

The classification tree was implemented as a Cycling'74 Max patch for online recognition. The vocal input undergoes the same processing adopted for the offline training so each imitation is analyzed and then the Median and IQR values are used in the classification tree. The classifier selects one of the classes, and appropriate sound feedback is returned.

The robust classification of vocal imitations is an open research topic which is out of the scope of this paper. By using a simple tree classifier trained on standard audio features we aimed at testing the feasibility of people-centered classification and at understanding how a user would adapt to such classifier. In fact, as it is often found in interactive systems relying on machine recognition, the user adapts to the machine and learns the kinds of utterances that make model selection reliable. The user learning process, for a wider set of sound classes, may be similar to what was expected by the Graffiti handwriting recognition system for Palm OS. In Graffiti, a set of glyphs was derived as a simplification of handwritten letters, and a new user was supposed to adapt and simplify his personal writing. In the proposed realization with only five sound classes, thanks to the simplicity of the decision tree and to the possibility to interpret the decision branches, it is possible to make sense of user learning and adaptation in terms of audio features.

Individual-Centered Selection

The level of accuracy that automatic classification of vocal imitations may reach in the future is not known, but in general it will be limited by the variety of vocal mechanisms that the casual user may decide to use. An opposite approach is that of training a classifier by examples provided by a specific user. We tested this approach by using `MuBu` objects for content-based real-time interactive audio processing [25], under Cycling'74 Max (see Figure 2).

The `MuBu.gmm` object extracts Mel-Frequency Cepstral Coefficients (MFCC) from audio examples (at least one example per sound class), and models each sound class as a mixture of Gaussian distributions. In the recognition phase, the likelihood for each class is estimated, and it can be used as a mixing weight for the corresponding sound model.

The individual-centered selection of sound models, when included in miMic, implies a further sub-mode of use, namely the Train procedure. In our realization the Train procedure is activated when both the buttons of miMic are pushed. To avoid using the GUI for this personalization stage, the user is requested to produce one vocal imitation for each of the five

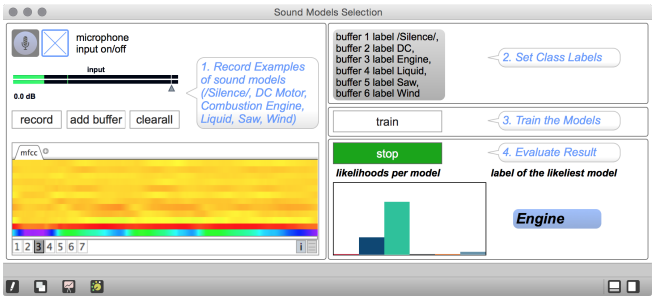


Figure 2. `MuBu.gmm` **used for individual-centered training.**

different sounds classes in a precise order, with an audible feedback marking the start and end of each imitation.

Feedback

The sound designer using miMic is supposed to face stages 0 to 2 of the sound design process (Table 1) without GUI support, free to use voice, gesture, two buttons, and active listening. Feedback for the user actions is essentially auditory, if we exclude the two LEDs that come as rings around the buttons. Feedback for stages 1 and 2 is the synthetic sound that the models are producing, when directly and continuously controlled by voice and gesture. Feedback to Select actions, on the other hand, is discrete, and two options have been explored: 1. Reference recording (e.g., a recording of water dripping as a response to a liquid vocal imitation); 2. on-the-fly synthesis (e.g., real-time synthesis of a liquid sound). The first gives immediate feedback about the recognized sound class, but gives no hint about the sonic possibilities of the sound model that will be used to represent it. That is why we introduced the sound model itself for selection feedback, with synthesis parameters tuned to two principal dimensions of vocal sound control: pitch and energy (whose periodic variations may elicit tempo perception). These are the dimensions that most people can reliably control and reproduce [17]. Each sound model has a control layer that is programmed to follow the pitch and amplitude contours of the provided imitation, so that the feedback, albeit being discrete and representing a sound class, mirrors the vocalizations used in the Select stage. For example, if a certain pitch profile is used to mimic a DC motor, the same profile will be used to vary the RPM model parameter (revolutions per minute) during playback.

Select or Mix?

A final, important question for the Select stage is whether the user might want to select a single model or a mixture of models in the set of available sound models. In a sense, it is like deciding if the user should, in a drawing program, choose from a set of predefined shapes (e.g., circle, rectangle, etc.) or if she would be allowed to combine a subset of those shapes (e.g., a rectangle with rounded corners). Generally speaking, playing with a single model is easier, but if the automatic classifier makes mistakes, and selects the wrong model, then the user might get frustrated by being unable to reach what she has in mind. Conversely, if the classifier provides

likelihoods, multiple models can be weighted and mixed together. This can be achieved thanks to the `MuBu` objects used for individual-centered classification which returns the likelihood of each class (see Figure 2).

PLAY

To demonstrate the Play mode of miMic we use a Cycling'74 Max patch that collects the sound models. For each of them, a control layer has been built and tuned in such a way that the vocalizations can be readily used as control signals. The assumption is that if a user selects a sound model such as combustion engine, then she will start controlling the model by producing engine-like sounds with the voice (stage 1, Mimic). Therefore, the control layer associated with the model must be ready to interpret such kind of control sounds mimicking the sound model. Only at a later stage the user might want to explore different vocal emissions (stage 2, Explore) and eventually to tune parameters by hand (stage 3, Refine), by specifying detailed maps between vocal features and model parameters on GUI. The Play mode offers a predefined, constrained, yet intuitive playground, whose boundaries are set by the affordances resulting by the combination of the sound model space with the control layer. However, stages 1 and 2 emphasize naturally the user's vocal and motor abilities, in using the tool at-hand. To stretch the 'pencil' metaphor further, stage 3 can be seen as the phase where users can 'sharpen' and perfect their tool, either to enhance or to support performativity.

A Set of Models

The five sound models, currently used in miMic's application, are a subset of the Sound Design Toolkit[9], a corpus of physics-based algorithms for the sound synthesis of machines and mechanical interactions. This collection provides a set of parametrized models that cover the taxonomy of perceptually-meaningful classes of vocal imitations, as devised in the project SkAT-VG. The goal is to provide sound designers with intuitive, procedural audio engines. In these models, sound is conceived as an acoustic behavior, resulting from the computed description of physical processes, according to configurations of simplified, yet perceptually-relevant sets of parameters [10].

Control Layers and Mapping

The basic control layer that each sound model is provided with is aimed at letting the user explore the available sonic space by exploiting a wide range of vocal and manual gestures. In the construction of the control layer, we must consider the limits of humans in controlling the dimensions of timbre with their voice [17]. Humans can reproduce pitch and tempo quite accurately, but they are far less consistent for timbre attributes such as sharpness. Therefore, we designed the basic control layer in order to map by default pitch and energy features to the most perceptually-salient parameters of each sound model. This control-layer design is largely based on deep knowledge of the underlying sound model, and on the phenomenological description of its parameters. Figure 3 shows how pitch is mapped to RPM in the combustion engine model. While the action of one control parameter (vocal or gestural feature) on one or many synthesis parameters (one-to-many mapping) is readily specified with the provided control layer, many-to-one maps may be necessary to fuse various control streams. One example may be the freezing of a pitch estimate when a low value of pitch clarity is returned by the pitch detector. Doing such operation within the visual language of Cycling'74 Max is daunting, but a simple JavaScript external (e.g., function `pitchclarity` in Figure 3) does the job.

[9] https://github.com/SkAT-VG/SDT

GESTICULATE

Humans often accompany vocal imitations with manual gestures, which may mimic the sound production mechanism or communicate some aspects of the sound morphology [6]. Thanks to its embedded IMU, miMic is designed to exploit either kind of gestures in sound sketching. The exploitation in the Select mode is problematic, as gestures are found to be highly variable, especially when they are of the mimicking kind. Conversely, some relevant dynamic behaviors, or movement qualities [1], can be extracted by processing the low-level signals returned by the IMU. In particular, recent research [26] has shown that humans consistently use shaky gestures to highlight the noisy quality of sounds they are imitating.

In the current realization of miMic the following movement qualities are extracted as continuous values: energy; shake; spin. Additionally a "kick detector" raises a flag in response to rapid acceleration changes. The dynamic behaviors, represented by continuously varying signals, are mapped to sound synthesis parameters to reinforce or complement the control actions exerted by the voice. One simple such mapping is preset and included in the control layer of each sound model.

In stage 2 (Explore) of the sound design process, the user typically realizes that the limitations in the simultaneous voice control of several synthesis parameters [17] can be partially overcome by using manual gestures. For example, the user may focus on voice pitch and roughness as control dimensions of timbre, while using gestural strokes to impart a bold temporal envelope. Something similar is commonly done by professional singers who move relatively to the microphone to hide a lack of sustained breath control. In general, synchronous vocal and motor control can be used to magnify a fine control on the sonic dimension. However, the development of interdependence between the vocal muscles and the limbs progresses with training and learning. This opens wide spaces for virtuoso performance and a careful moulding of the desired sound, by playing miMic, and programming and tuning arbitrarily complex maps at stage 3 (Refine).

A simple sensor fusion strategy based on JavaScript externals is used to support such shifts of control from voice to gesture (e.g., function `passthru` in Figure 3).

DEVELOPMENTS

The proposed system architecture is aimed at supporting ubiquitous sketching in sound design practices. This design goal will drive future improvements and developments.

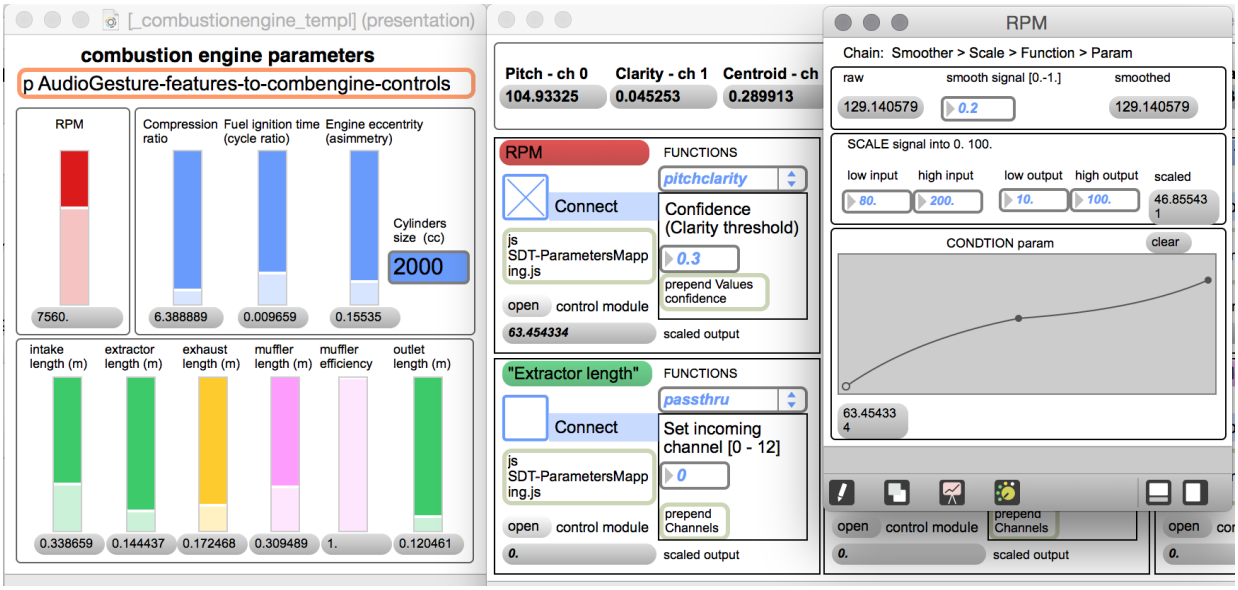


Figure 3. Fine control of the parameters of a combustion engine sound model: Nonlinear control of RPM by pitch.

In the current realization, each sound model is equipped with a simple control layer that maps audio and gesture features into model parameters, and the sound designer is free to introduce additional control complexity at stage 3 (Refine) of the design process. One other possibility would be to perform "mapping by demonstration" [11], i.e., to ask the user to perform exemplary gestures to accompany a set of given sound examples, so that the system may be able to generalize and respond to other gestures. Such a feature may be introduced in future extensions of miMic.

In the short term, miMic is going to integrate all the sound categories studied in the SkAT-VG project, as well as the general imitation classifier that the project is producing. miMic will be used as an experimental tool in sound design workshops, and the merits and pitfalls of the architecture will be experimentally evaluated, especially regarding the fluidity and effectiveness of individual and collaborative sketching workflows.

ACKNOWLEDGMENTS

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

REFERENCES

1. Sarah Fdili Alaoui, Baptiste Caramiaux, Marcos Serrano, and Frédéric Bevilacqua. 2012. Movement Qualities As Interaction Modality. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 761–769. DOI: http://dx.doi.org/10.1145/2317956.2318071
2. David Sanchez Blancas and Jordi Janer. 2014. Sound retrieval from voice imitation queries in collaborative databases. In *Audio Engineering Society Conference: Semantic Audio*. Audio Engineering Society.
3. Bill Buxton. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
4. Serena Cangiano and Davide Fornari. 2014. Products As Platforms: A Framework for Designing Open Source Interactive Artifacts. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems (DIS Companion '14)*. ACM, New York, NY, USA, 219–222. DOI: http://dx.doi.org/10.1145/2598784.2598805
5. Baptiste Caramiaux, Alessandro Altavilla, Scott G. Pobiner, and Atau Tanaka. 2015. Form Follows Sound: Designing Interactions from Sonic Memories. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3943–3952. DOI: http://dx.doi.org/10.1145/2702123.2702515
6. B. Caramiaux, F. Bevilacqua, T. Bianco, N. Schnell, O. Houix, and P. Susini. 2014. The Role of Sound Source Perception in Gestural Sound Description. *ACM Trans. Appl. Percept.* 11, 1, Article 1 (April 2014), 19 pages. DOI: http://dx.doi.org/10.1145/2536811
7. Mark Cartwright and Bryan Pardo. 2015. VocalSketch: Vocally Imitating Audio Concepts. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 43–46. DOI: http://dx.doi.org/10.1145/2702123.2702387

8. Stefano Delle Monache, Davide Rocchesso, Stefano Baldan, and Davide Andrea Mauro. 2015. Growing the practice of vocal sketching. In *21st Proc. International Conference on Auditory Display*. Graz, Austria, 58–65.
9. Inger Ekman and Michal Rinott. 2010. Using Vocal Sketching for Designing Sonic Interactions. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*. ACM, New York, NY, USA, 123–131. DOI: http://dx.doi.org/10.1145/1858171.1858195
10. Andy Farnell. 2010. Behaviour, structure and causality in procedural audio. In *Game sound technology and player interaction concepts and developments*, M. Grimshaw (Ed.). Information Science Reference, New York, NY, USA, 313–329.
11. Jules Françoise. 2013. Gesture–sound Mapping by Demonstration in Interactive Music Systems. In *Proceedings of the 21st ACM International Conference on Multimedia (MM '13)*. ACM, New York, NY, USA, 1051–1054. DOI: http://dx.doi.org/10.1145/2502081.2502214
12. Gabriela Goldschmidt. 1991. The dialectics of sketching. *Creativity Research Journal* 4, 2 (1991), 123–143. DOI: http://dx.doi.org/10.1080/10400419109534381
13. Saul Greenberg, Sheelagh Carpendale, Nicolai Marquardt, and Bill Buxton. 2011. *Sketching User Experiences: The Workbook* (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
14. Donna Hewitt and Ian Stevenson. 2003. E-mic: extended mic-stand interface controller. In *Proceedings of the 2003 conference on New interfaces for musical expression*. National University of Singapore, 122–128.
15. Daniel Hug and Moritz Kemper. 2014. From Foley to function: A pedagogical approach to sound design for novel interactions. *Journal of Sonic Studies* 6, 1 (2014). http://journal.sonicstudies.org/vol06/nr01/a03
16. Jordi Janer and Maarten De Boer. 2008. Extending voice-driven synthesis to audio mosaicing. In *Proc. Sound and Music Computing Conference*. Berlin, Germany.
17. Guillaume Lemaitre, Ali Jabbari, Olivier Houix, Nicolas Misdariis, and Patrick Susini. 2015. Vocal imitations of basic auditory features. *The Journal of the Acoustical Society of America* 137, 4 (2015), 2268–2268. DOI: http://dx.doi.org/10.1121/1.4920282
18. Guillaume Lemaitre and Davide Rocchesso. 2014. On the effectiveness of vocal imitations and verbal descriptions of sounds. *The Journal of the Acoustical Society of America* 135, 2 (2014), 862–873.
19. Enrico Marchetto and Geoffroy Peeters. 2015. A set of audio features for the morphological description of vocal imitations. In *Proc. Int. Conf. on Digital Audio Effects*. Trondheim, Norway.
20. Stefano Delle Monache, Stefano Baldan, Davide A. Mauro, and Davide Rocchesso. 2014. A design exploration on the effectiveness of vocal imitations. In *Proceedings Intern. Computer Music Conf. / Conf. on Sound and Music Computing*. Athens, Greece, 1642–1648.
21. Fred Newman. 2004. *MouthSounds: How to Whistle, Pop, Boing and Honk for All Occasions... and Then Some*. Workman Publishing Company, New York.
22. Elif Özcan, René van Egmond, and Jan J. Jacobs. 2014. Product Sounds: Basic Concepts and Categories. *International Journal of Design* 8, 3 (2014), 97–111. http://www.ijdesign.org/ojs/index.php/IJDesign/article/view/1377
23. Yongki Park, Hoon Heo, and Kyogu Lee. 2012. Voicon: An Interactive Gestural Microphone For Vocal Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, G. Essl, B. Gillespie, M. Gurevich, and S. O'Modhrain (Eds.). University of Michigan, Ann Arbor, Michigan.
24. Dan Moses Schlessinger. 2012. Concept Tahoe: Microphone Midi Control. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, G. Essl, B. Gillespie, M. Gurevich, and S. O'Modhrain (Eds.). University of Michigan, Ann Arbor, Michigan.
25. Norbert Schnell, Axel Röbel, Diemo Schwarz, Geoffroy Peeters, Riccardo Borghesi, and others. 2009. MuBu and friends–Assembling tools for content based real-time interactive audio processing in Max/MSP. In *Proc. International Computer Music Conference*. Montreal, Canada.
26. Hugo Scurto, Guillaume Lemaitre, Jules Franoise, Frdric Voisin, Frdric Bevilacqua, and Patrick Susini. 2015. Combining gestures and vocalizations to imitate sounds. *The Journal of the Acoustical Society of America* 138, 3 (2015), 1780–1780. DOI: http://dx.doi.org/10.1121/1.4933639
27. D. Stowell. 2010. *Making music through real-time voice timbre analysis: machine learning and timbral control*. Ph.D. Dissertation. School of Electronic Engineering and Computer Science, Queen Mary University of London. http://www.mcld.co.uk/thesis/
28. Tiffany Tseng. 2016. Build in Progress. In *Makeology: The Maker Movement and the Future of Learning*, K. Peppler, E. Halverson, and Y. Kafai (Eds.). Vol. 2. Routledge, New York, NY. In preparation.
29. Benjamin Vigoda and David Merrill. 2007. JamiOki-PureJoy: A Game Engine and Instrument for Electronically-mediated Musical Improvisation. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression (NIME '07)*. ACM, New York, NY, USA, 321–326. DOI: http://dx.doi.org/10.1145/1279740.1279810