Davide Rocchesso, Iuav University of Venice
Guillaume Lemaitre, IRCAM
Patrick Susini, IRCAM
Sten Ternström, KTH
Patrick Boussard, Genesis

Sketching Sound with Voice and Gesture

Insights

→ In product and interaction design, sounds should be included in the early stages of the design process.

→ Voice and gestures are natural sketching tools that we can exploit to communicate sonic interactions.

What does a sound designer do? Walter Murch is credited for introducing the term *sound designer* both to describe his own contribution to the movie *Apocalypse Now* and to circumvent the constraints that trade unions imposed on the film industry, because his job was neither sound editor nor sound director. George Lucas, talking about the science fiction movie *THX1138*, used the following words to describe his work with Murch: "We took the sound effects and made them to be like music, and in some cases, we took the music and made it to be sound effects." In a sense, aural form was following function, and occasionally aural function was following form [1].

In the past half-century, the term *sound design* has been used in a variety of contexts, ranging from the arts to industrial products. Only the emergence of interactive uses of sound, however, has pushed many researchers to consider the sound studies from a design perspective. Sonic interaction design is now an active area of research whose methods and tools are being exploited in design practices as well as in other scientific fields [2]. Sound is no longer considered in isolation, looked at only as an object to be analyzed, processed, or edited. Rather, investigators direct their attention toward the aural manifestations of interactive objects. It is in interactive contexts that the study of the meaning of a sound can be considered in terms of an action-

perception process.

Neuroscientists have found that there are audiovisual mirror neurons in the monkey premotor cortex that discharge when the animal performs, sees, or hears a specific action. Scientists of human motion have shown that auditory stimuli are important in the performance of difficult tasks and can, for example, elicit anticipatory postural adjustments in athletes. These findings justify the attention given to sound in interaction design for gaming, especially in action and sports games that afford the development of levels of virtuosity. But there are many everyday situations when a careful sound design would positively affect efficiency, fluency, or safety of interactions while reducing the sonic clutter that we experience in many contemporary environments.

The sonic manifestations of objects can be designed, both by acting on their mechanical qualities and by augmenting the objects with synthetic, responsive sounds. Many new products have no mechanical moving parts; in these cases, the introduction of electronic noises is often a necessity [3]. This is where designing sound ultimately means devising models for "procedural audio" [4] that respond to actions or environmental changes. Although a large repertory of synthesis models and software environments are readily available, the work of sonic interaction designers is impaired by the lack of effective sketching tools, thus making integration with existing interaction design practices difficult.

Actually, we are all naturally equipped with a phenomenal sound sketching tool—our voice—that we often use to mimic and communicate everyday sounds. To communicate sonic interactions, we can accompany vocal imitations with gestures. We often repress our vocal and gestural expression due to social conventions, but just observing a child playing with a toy car gives us the feeling of how powerful voice and gesture are as sketching tools. Adults also spontaneously, and often subconsciously, use vocalizations and gestures to communicate sounds. In order to become usable in contemporary interactive applications, however, the

sonic and gestural sketches should be captured and converted into procedural audio, or into sound models whose parameters can be tweaked and used to actually design sound in action. The EU project SkAT-VG (Sketching Audio Technologies using Vocalizations and Gestures; http://www.skatvg.eu/) aims to bridge the gap between utterances and sound models so that designers can start using their voice and gestures as naturally and fluidly as when drawing with a pencil.

SOUND DESIGN

Sound is an important component of a wide range of products and contexts, from games to vehicles, from home appliances to public spaces. The professionals involved in designing these sounds can be engaged in very different activities, from audio sample selection to modification of physical mechanisms. In many applications, sounds contribute to the overall experience of a product, for example by conveying a sense of solidity or power. In general, sounds play an important role in the emotional and aesthetic qualities of products.

Today, an application area that is particularly interesting for sound design is that of electric or hybrid cars, where the driving experience emerges from complex continuous interactions. Real-time audio processing and even sound synthesis can be used to give the driver a rewarding experience. But even more important for the safety of pedestrians is the sound that cars produce when running at low speeds, when tire and wind noise can go unnoticed against a loud background. In this case, if the engine is inherently silent, the sound designer has a vast palette of possibilities and must devise a dynamic sound model that conveys the required information. The adjective *dynamic* here means that sounds should be procedurally generated to give the pedestrian continuous and reliable cues about the speed of the approaching vehicle. These sounds cannot be just discrete alarms; they should respond continuously and expressively to the control actions of the driver.

Moreover, these newly designed sounds should contribute to embellishing our environment by creating a pleasant new soundscape, as was already considered in the 1970s by Murray Schafer, the father of soundscape studies: "In the new soundscape, one must worry about the prevention of sounds as well as their creation." Achieving optimal results may require extensive and accurate testing at all stages of the design process. As an example, Figure 1 shows a picture that was shot during a sound design session for electric vehicles at Genesis. But it is at the early stages of the sound design process that the designer needs to be able to quickly produce, compare, and evaluate ideas of sonic interaction—in other words, to sketch sonic behaviors.

VOCALIZATIONS AND GESTURES

If a picture can be worth a thousand words, how much is a vocalization worth? Humans use only a subset of their vocal possibilities to utter concepts and sentiments in the form of words. Conversely, the whole range of vocal possibilities is available for imitating sounds of objects or animals, or simply to communicate sonic ideas.

Recent research has shown that listeners recognize the categories of sounds that have been vocally imitated. Moreover, when asked to communicate a given sound, humans are even more effective with imitations than with verbal descriptions [5], just as a drawing can more clearly communicate a concept than can a written sentence. This is not to diminish the power of specialized language and verbal descriptions, as it has also been shown that perceptually relevant descriptions are effective for identifying a target sound in a set.

Current research is trying to parse the processes that subserve such tremendous performance and to identify how articulatory mechanisms and the acoustic properties of the imitated sounds are related. It is also interesting to see what happens when hand gestures are used to communicate sounds. It turns out that humans prefer to mimic the sound-producing action if they are able to identify the sound source ("mimicking"), otherwise reverting to a gestural replication of spectro-morphological features of the sound ("tracing") [6]. In any case, voice and gesture provide a rich domain for sketching that is just waiting for the appropriate tools that can be exploited for sonic interaction design.

SKETCHING SOUND

In an article for *Interactions*, Paul Robare and Jodi Forlizzi encouraged designers to "begin considering the sounds of their products at the concept-generation stage" [7]. Unlike their colleagues in other areas of design, however, sound designers are not accustomed to teamwork and participatory practices. They often refrain from sharing preliminary sketches, and they prefer to offer highly refined realizations to their clients. In this sense, sound design seems to be closer to art—sometimes to black art—than to design. However, as soon as sound and interaction become tightly intertwined in interactive products, the need to embrace a well-formed design process becomes evident.

In the SkAT-VG project, as well as in prior research [2], we have been promoting research through design, where raw models and sketches of sonic interactive scenarios and artifacts become means of embodied design thinking. The workshop is the natural venue where shared doing, reflective practices, and inter-observation enable understanding through designing interactive sonic objects. It is in these kinds of activities that voice and gesture express their effectiveness as sketching tools for sonic interaction design. In particular, the combined use of manipulation of physical props (Foley artistry), vocalization, and sound synthesis proved effective for the rapid generation and evaluation of sonic interactions. In a workshop setting, the different skills and backgrounds of participants offer good opportunities to improve the variety and quality of sketches. Much more easily than in visual sketching, participants can be made to cooperate to produce joint sketches that

Voice and gesture provide a rich domain for sketching that is just waiting for the appropriate tools that can be exploited for sonic interaction design.

Figure 1. Sound designer Andrea Cera testing prototype sounds for an electric car.

ENDNOTES

1. Pauletto, S., ed. *The New Soundtrack, special issue on Perspectives on Sound Design 4*, 2. Edinburgh University Press, 2014.
2. Franinović, K. and Serafin, S., eds. *Sonic Interaction Design*. The MIT Press, Cambridge, MA, 2013.
3. A classic example is the Mighty Mouse introduced by Apple in 2005, where the scrolling sound was produced by a tiny speaker inside the mouse.
4. Farnell, A. *Designing Sound*. The MIT Press, Cambridge, MA, 2010.
5. Lemaitre , G. and Rocchesso, D. On the effectiveness of vocal imitations and verbal descriptions of sounds. *J. Acoustical Society of America 135*, 2 (2014), 862–873.
6. Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O., and Susini, P. The role of sound source perception in gestural sound description. *ACM Trans. Appl. Percept. 11*, 1 (2014).
7. Robare, P. and Forlizzi, J. Sound in computing: A short history. *Interactions 16*, 1 (2009), 62–65.

⬤ **Davide Rocchesso** is an associate professor at the Iuav University of Venice, where he coordinates the SkAT-VG project. From 2007 to 2011, he chaired the COST Action IC-0601 SID (Sonic Interaction Design).
→ roc@iuav.it

⬤ **Guillaume Lemaitre** is a research associate with IRCAM. His work focuses on auditory cognition and how listeners process auditory signals to infer information about the sources of the sounds and interact with them. His research is applied to product sound quality and sonic interaction design.
→ guillaumejlemaitre@gmail.com

⬤ **Patrick Susini** is the head of the Sound Perception and Design team at IRCAM. He received a Habilitation in 2011. He has been teaching in the Master of Architectural Acoustics program (Paris 6) since 1997, and in the Master of Sound Design (ESBAM) program since 2010.
→ patrick.susini@ircam.fr

⬤ **Sten Ternström** is a professor at KTH Royal Institute of Technology in Stockholm, where he also coordinates the FET-Open EUNISON project. From 2007 to 2010, he was a Swedish delegate to the COST Action 2103 (Advanced Voice Function Assessment). His research interests include voice analysis and synthesis, as well as engineering aspects of music and speech.
→ stern@kth.se

⬤ **Patrick Boussard** is the executive director and founder of Genesis, a company based in Aix-en-Provence, France. The company focuses on sound simulation, sound perception, and sound design for the automotive and the aeronautics industries.
→ patrick.boussard@genesis.fr

overcome some of the limitations of the human voice, for example, to produce polyrhythmic or multipitched collective vocalizations. Almost inevitably, these sketching activities turn into theatrical performances very similar to what other areas of interaction design have been experimenting with through bodystorming.

CHALLENGES

A move to a new sound design tool!

If the virtues of voice and gesture are so evident, what is missing to make them the principal sketching tools for sound designers? The answer to such a question is multifaceted. In visual and product design, the current practices have been heavily shaped by a century of pedagogical and research experiences that have been gradually separating design from the arts. In sound design, such a cultural path has not been followed with the same determination, and sound design practices have been established and widely shared only in cinema (Foley) or in activities close to music (sound art). As a consequence, the importance of sketching with sound has not yet been fully expressed, and the awareness of what we can use and do at the initial stage of a sound design process is not widespread. Learning to use the sketching possibilities of our voices is certainly a necessary first step, but major progress probably will be achieved when technology becomes available to support the designer in the rapid production of sonic sketches.

Ideally, it would be desirable to convert vocal and gestural sketches into models of sound production, whose parameters can be further tuned and controlled in the context of use. Such tuning and control could also be done using voice and gesture, for the sake of immediacy, naturalness, and expression. In other words, voice and gesture would provide both the sonic material and the manipulative processes that are necessary to convert sketches into prototypes. Future sketching tools will realign design thinking and crafting, thus recovering the embodied and performative qualities of interaction design practices. This will help interaction designers conceive and develop products that are genuinely multisensory, in their look and feel as well as in their embodied functionality.

These are the final objectives of the SkAT-VG project, which will be eventually achieved after a fair amount of research in the areas of sound perception (vocal imitations), sound analysis (feature extraction), machine learning (classification), and sound synthesis (parameter mapping and control). Keep an ear out for new results!

ACKNOWLEDGMENTS