SONIC GESTURES AS INPUT IN HUMAN-COMPUTER INTERACTION: TOWARDS A SYSTEMATIC APPROACH

Antti Jylhä
Department of Signal Processing and Acoustics
Aalto University, School of Electrical Engineering
antti.jylha@aalto.fi

ABSTRACT

While the majority of studies in sonic interaction design (SID) focuses on sound as the output modality of an interactive system, the broad scope of SID includes also the use of sound as an input modality. Sonic gestures can be defined as sound-producing actions generated by a human in order to convey information. Their use as input in computational systems has been studied in several isolated contexts, however a systematic approach to their utility is lacking. In this study, the focus is on general sonic gestures, rather than exclusively focusing on musical ones. Exemplary interactive systems applying sonic gestures are reviewed, and based on previous studies on gesture, the first steps towards a systematic framework of sonic gestures are presented. Here, sonic gestures are studied from the perspectives of typology, morphology, interaction affordances, and mapping. The informational richness of the acoustic properties of sonic gestures is highlighted.

1. INTRODUCTION

The connection between gesture and sound has been a point of intensive study during the past decade in the field of sound and music computing (SMC). We have witnessed numerous musical performances, software applications, and human-computer interfaces highlighting the use of gestural control of sound output. Also theoretical advances in action-sound relationship and sound-related gesture research have been presented, making use of both results and theories from the field of human-computer interaction (HCI) and classical theories on sound objects, and applying these to SMC especially in the domain of gesture in music [1, 2, 3, 4, 5].

The mainstream of gesture research in the field of SMC has concentrated on ”traditional” viewpoints of gestures and their acquisition. This is to say that although definitions of gesture differ dependent on the context, gesture is seen as a human-generated perceivable physical action, which is often analyzed by means of haptic input or computer vision. For example, in the context of sonic interaction design (SID), most of the presented interaction paradigms and interfaces rely on haptic and/or visual input, which is analyzed to inform sound production. In musical controllers, it is usually the physical gesture that is mapped to the sound output, either as sound-generating or sound-modifying action.

However, SID has been defined as studying sound as the conveyor of information, aesthetics, and emotions, which does not imply a one-way approach to the use of sound at the interface. Instead, it can be interpreted to also consider sound propagating to the other direction in the interaction loop, i.e., from the human to the computer. Only recently the sounds generated by humans have started to gain more attention in the field in the context of controlling interfaces and applications. Looking back, isolated examples and studies on sonic input can be found, but a common perspective on using sound as a key input modality has not been presented.

This study concentrates on the notion of sonic gesture as input in HCI, which is an interesting topic for several practical reasons. First, sounds do not require specialized hardware to acquire, as most computational devices, be they computers or mobile devices, are equipped with a microphone. Rather, the challenge is in the processing of the sounds to acquire meaningful information from them [2]. Second, sonic gestures facilitate remote interactions, i.e., the user does not need to touch the operated device. Third, sound as an input modality can work in situations, where looking at the device is not possible (eyes-busy situations, visual impairment). Fourth, some sonic gestures can provide an alternative means of accessing computers and applications for people with motor impairment.

This work will discuss the use of sonic gestures as input in HCI and SMC fields, basing the discussion on previous studies and examples of sonic gestures in action. While a large portion of previous studies are musically oriented, this study considers sonic gestures in general, stripped from the constraint of their use exclusively in musical contexts. The aim is to take steps towards a systematic approach to sonic gestures in terms of the types of interaction and information different sonic gestures afford.

2. SONIC GESTURES

This work defines sonic gesture as a sound-producing action generated by a human in order to convey information to a computational system. This definition differs from previous definitions in that the gesture itself is always a sound producing action and does not necessitate an instrument for production, although sonic gestures can be instrumental, too. Furthermore, while there is a vast and growing body of research looking into gesture from the musical perspective, in this study the sonic gesture itself does not necessarily comprise musical elements.

Considering the above definition, it is important to note that the perspective of this study differs from the classical action-sound perspective. Here, the sound always occurs prior to computation, i.e., this work does not consider as sonic gesture for example wielding an accelerometer-equipped controller in the air and mapping this motion into sound synthesis parameters. In this work, the information conveyed lies within the human-generated sound itself, and the examination focuses on its acoustic properties and the use of these for informing or controlling an interactive system, rather than trying to infer the gesture behind the sound.

While higher-level aspects of sonic gestures, such as emotion, social connotations, or Chion's concept of *ergo-audition* [6] (the experience of hearing the sounds of one's own acting) are definitely relevant for utilizing sonic gestures in context, from HCI perspective capturing this information from sound with a computational system is still mostly a challenge of the future. Nevertheless, when sonic gestures are performed in interactive contexts, there are actually two levels of feedback the user gets: the sound and sensation from the sonic gesture itself, and the feedback from the computational system. The former is in practice always multisensory, containing typically auditory and haptic components, while the modalities of the latter vary dependent on the form, function, and design of the interactive system.

The simple case of a hand clap is a good basic example of a sonic gesture. It is clear that it is a sound-producing action - with a very distinct sound - generated by a human. As will be discussed below, this simple gesture can convey lots of information with its acoustic properties. While a hand clap can be considered a gesture by itself, the information it conveys is ultimately dependent on the context in which it is produced. It can be argued that sonic gestures become meaningful only when they have been associated with a meaning, which in HCI is achieved usually by means of mapping the gesture or some of its properties to a command on the computational device.

Jensenius has categorized definitions of gesture into three groups: gesture as communication, gesture for control, and gesture as mental imagery [3]. In communication, gesture is seen as means for social, interpersonal interaction, whereas mental imagery refers to studying gestures as mental processes. In this study, the focus is mainly on gestures for control, which can be seen as the traditional HCI perspective for gestures. However, as we shall see from examples, some approaches to sonic gesture also relate closely to both communication and mental imagery, even if the aim is in control. Also, control in this work refers not only to giving commands but to convey information that is essential for interaction in a broader context.

In contrast to the gesture definition of Cadoz [7], which excludes all vocal sounds, sonic gestures can be produced also vocally. Indeed, non-speech utterances, humming, and mouth-generated sounds provide a very rich gestural repertoire. Their utility in designing sonic interactions has been discussed by Ekman and Rinott [8] in their influential work on vocal sketching. Dessein and Lemaitre [9] have explored the capability of humans to imitate everyday sounds vocally, and found out that there exists a strong connection between the classification performed by humans and that performed for real everyday sounds by acoustic descriptors. In addition, as shown by Sporka [10], there is a lot of intuitive information in how people communicate by pitch alone to indicate confirmation, negation, uncertainty, and surprise, among others. While it is a very simple acoustic descriptor, pitch can be utilized in numerous ways in interactive systems, too.

Van Nort has studied sonic gestures from a musical perspective [4]. He defines sonic gestures in the context of instrumental excitation and interaction design, and carefully dissects the gestures into possible control structures based on their acoustic morphology. He presents a perspective on mapping as more than just the link between control and output, highlighting the importance of mental images in the perception on musical dynamics and the gestures used in music production.

2.1 Previous studies on sonic gesture interfaces

The range of sonic gestures is broad and several studies have proposed interfaces using some type of sonic gesture for a particular problem. While the examples here are by no means exhaustive, they show the variety of different gestures and approaches to their utilization, which will be used as basis for discussion in Section 4.

Vesa and Lokki have presented a music player control interface using finger snaps [11]. The system utilizes two microphones integrated to the headphones of the user and is capable of detecting which snaps occur on the left or right side of the user's head or in front of it. This information is mapped to previous/next track and play/pause functionalities found in all music players nowadays.

Jylhä and Erkut have developed a hand clap interface for sonic interactions with the computer [12]. From a stream of percussive sound events, the interface can extract information on the event type (i.e., hand configuration) and tempo. This information can then be used to indicate control information in various applications. It has been demonstrated on giving discrete commands to the system, controlling the tempo of music, and entraining a virtual audience to the user's clapping. More recently, the interface has been applied and extended to an interactive Flamenco hand clapping tutor application, in which also accentuation (clap strength) and temporal deviation of the user's clapping are extracted [13]. This information is applied to inform rhythmic output from the system, and to monitor the performance of a learning clapper.

Rocchesso, Polotti, and Delle Monache have studied continuous sonic interactions based on kitchen activities and sounds [14]. As one case example, they consider carrot cutting, a rhythmic activity, which they sonify with several different feedback strategies. As one input modality, they utilize the contact sound resulting from the knife hitting the table, and perform beat-tracking on the sound. Providing rhythmic sonic feedback with an adaptive tempo and upbeat rhythm seemed to result in the most relaxed action by the cutter.

Vocal Joystick [15] is an interface enabling the user to control for example the mouse cursor by vowel sounds. From vowels, the interface extracts energy, pitch, and vowel quality. Energy is mapped to the velocity of cursor movement, while vowel quality is mapped with a continuous two-dimensional mapping into movement direction. The Vocal Joystick has been shown to compete with eye-tracking based cursor movement interfaces.

Another mouse-replacement interface has been presented in [16], based on humming and hissing. A four cell mouse grid is used, and a cell is selected by low-frequency or high-frequency humming. Hissing is detected and mapped to a mouse click event.

Sporka [10] has presented several methods and applications around using pitch-based vocal input in HCI, including target acquisition by absolute and relative pitch, mouse cursor control by whistling or humming using pitch and loudness parameters, non-speech control of keyboard emulation by mapping three-element pitch patterns to keyboard keys thus forming an alphabet of sonic gestures, and controlling two computer games by vocal input. The methods have been designed especially for hands-busy situations and people with motor impairment.

Hämäläinen has presented computer game applications incorporating vocal input as part of the interface [17]. In one game, shouting is used to control the fire-breathing of a dragon avatar, while in other ones voice pitch controls the avatars' movements.

Billaboop is an interface which allows the user to play virtual drums by beatboxing sounds [18]. The system captures the sonic gestures of the user and by means of machine learning triggers drum sounds corresponding to the detected sounds. The system is also capable of reacting to table drumming in the same way. Recently, a mobile application called BoomClap [1] from the same origin has been presented. The application can be taught which sounds the user wants to use in the interaction.

Considering instrumental sonic gestures, Scratch Input demonstrates how scratches on surfaces can form a rich sonic gesture repertoire [19]. The sounds are captured by a contact microphone and different gestures are recognized by their sound trajectory. Given that structure-borne sound travels far and is highly tolerant for unwanted environmental sounds, the technique is relatively robust.

As an innovation to musical applications for mobile devices, the iPhone Ocarina application presents an interface where the blowing of the user to the microphone of the device acts as excitation of sound [20]. This is a good example of a natural interface, where the interaction with the computational device is very close to that with a real ocarina.

3. TYPOLOGICAL AND MORPHOLOGICAL CONSIDERATIONS

As discussed for example by [5] and [4], gestures can be divided into three categories based on their macro-level morphology: impulsive, iterative, and sustained gestures. Iterative and sustained gestures have also been labeled continuous gestures [1], but as argued by Van Nort, iterative and continuous gestures can be seen as separate categories [4]. Isolated hand claps and finger snaps, for example, are impulsive sonic gestures, whereas whistling and humming are sustained. Continuous hand clapping with a relatively constant tempo is an iterative gesture, consisting of sequential impulsive gestures. It is noteworthy, however, that in principle any basic gesture type - be it impulsive or sustained as an isolated case - can be sequentially produced to form iterative gestures.

Typologically, its is clear that a hand clap is a different type of gesture than a whistle. However, it is possible to expand the gestural typology by considering different types of the same gesture class as their own subtypes. For example, it has been shown that different hand configurations in clapping result in audibly different sounds and that it is possible by machine learning techniques to also differentiate between these types computationally [21, 22]. Furthermore, it is possible to consider such a typology as a continuum, as is done for example in the Vocal Joystick example for vowels [15]. While this continuum may rely on anchors, for example eight vowels of spoken language, to form a basis for the mapping space, the "in-between" vowels can be used to provide a continuum rather than a class-based typology. Similarly for pitch, it is possible to produce melodies by sounds of different constant pitches following each other, or to continuously vary the pitch, e.g., with a glide up and down in pitch.

Considering acoustic morphology further, sonic gestures can be categorized in several ways. There are unpitched sonic gestures such as hand claps, finger snaps, and table taps, and pitched ones like whistling and humming. On the other hand, the shape of sonic gestures can be static, e.g., humming with a constant pitch, or dynamic, e.g., humming with a varying pitch. It is possible to dwell deeper into the acoustic morphology, looking at Schaefferian principles of sound objects, as has been discussed by Van Nort [4] in the context of musical gesture, which would apply for the most part also to the morphologies of general sonic gestures. However, in this study the focus remains more on a macro-level.

It is also possible to see a connection between the sonic gesture typology and Gaver's map of everyday sounds [23], which examines the sounds generated by interacting materials starting from their fundamental sources (solids, gasses, liquids), and proceeding through basic sound-producing events into temporal patterning and more complex sounds. A similar approach can also be envisioned for sonic gestures, grouping them based on the basic-level events, temporal patterning, and combinatory events of multiple gestures.

An important aspect to consider in sonic gesture discussion is the sound-producing body. For sound-producing

[1] http://billaboop.com/en/boomclap

gestures, a categorization into empty-handed and instrumental gestures has been proposed [2]. Ballas [24] has labeled both of these as self-produced sound, including all sounds produced by the body or body movements, with or without interacting with an external surface or object. In the context of sonic gestures, empty-handed gestures can be considered as all sound-producing actions that the human is able to produce without an external sounding body. For example hand claps, finger snaps, whistling, all vocal sounds, and body tap sounds can be considered empty-handed gestures, even if hands can be used in their production. Instrumental sonic gestures are actions, in which the human interacts with a secondary physical object to generate sounds, e.g., footsteps and scratches or knocks on surfaces. Here, we consider as first-order instrumental sonic gestures those sound-producing actions, that involve direct human interaction on a secondary sounding body. Tapping a table or scratching a wall are first-order sonic gestures, as is blowing into the microphone of a mobile device as in [20] to create turbulent air flow sound. Second-order instrumental sonic gestures involve interacting with a secondary sounding body through a proxy object, for example hitting a table with a pen or throwing a ball to a wall. This class is so vast, however, that it is not discussed in this study.

4. EXTRACTABLE MAPPING PARAMETERS OF DIFFERENT SONIC GESTURES

Designing an interface around sonic gestures the designer is faced with the questions of what parameters of the sound need to be computed and how they can be mapped to the system functionality and output to provide meaningful interaction. Looking at the variety of sonic gestures, it is clear that different gestures can provide different types of information and, thus, are applicable for different purposes. Here we take a detailed look at an exemplary set of sonic gestures and present a set of parameters that can be computed from each gesture type, summarized in Table 1. The upper part of the table enlists empty-handed gestures, while the lower part considers first-order instrumental gestures. These parameters are non-exhaustive and relatively low-level, and it is in most cases possible to compute higher-level parameters as well.

In Table 1, every listed gesture has been categorized by the relevant temporal forms. As discussed above, it is noteworthy that any listed basic action can afford iterative gestures. Sequentially produced percussive gestures have been more widely used, but there is no reason why whistling or humming, for example, couldn't be produced iteratively as well. This also results in the fact that any gesture, when produced iteratively, can convey temporal parameters such as tempo, temporal deviation, acceleration slope etc. These continuous parameters can then be mapped to continuous commands and actions in the system. For example, monitoring the tempo of a clapping user can be used in a musical system to inform the tempo of the sound output as in [13]. As tempo is a continuous parameter, it could also be used to inform other than rhythmic functions in a system requiring continuous control. Considering richness of information, it should be noted that the iterative stream is still a result of concatenating basic gestures, and can incorporate informational parameters obtainable from the basic sounds, such as different hand configurations varying in the stream.

Sounding action (basic gesture)	Temporal form	Extractable parameters
hand clap	**impulsive**, **iterative**	**clap type**, patterns, **tempo**, **temporal deviation**, acceleration, **volume**
finger snap	**impulsive**, iterative	tempo, temporal deviation, acceleration, patterns
body tap	impulsive, iterative	type, tempo, temporal deviation, acceleration, patterns, volume
whistling	**sustained**, iterative	**pitch**, **duration**, **slope**, pattern, **volume**, tempo
vocal: humming etc.	**sustained**, iterative	**pitch**, **duration**, **slope**, **pattern**, **volume**, tempo, timbre
vocal: impulsive	**impulsive**, iterative	**type**, tempo, deviation, acceleration, patterns, volume
vocal: fricatives	sustained, iterative	type, duration, timbre, tempo, volume
vocal: vowels	**sustained**, iterative	**pitch**, **type**, **duration**, **timbre**, **volume**, tempo
breathing	sustained, iterative	type, duration, timbre, tempo, volume
footsteps	impulsive, iterative	type, tempo, patterns, volume
knocks and taps	**impulsive**, iterative	**type**, tempo, deviation, acceleration, patterns, volume
scratches	**sustained**, iterative	**type**, **shape**, duration, tempo, **patterns**, volume
blowing turbulence	**sustained**, iterative	**duration**, patterns, tempo, volume

Table 1. A set of empty-handed (top) and instrumental (bottom) sonic gestures with different morphologies, including the basic parameters that can be extracted from each gesture. **Bold** face signifies that the gesture, temporal form, or parameter has been utilized in one or more of the interfaces and applications summarized in section 2.1, while the rest of the items are considered feasible in practice as well.

For all temporal forms, an obvious piece of information lies in the very occurrence of the sound-producing event. As exemplified above and in [12], for example individual hand clapping sounds can be detected to give discrete commands or trigger actions in computational systems. Also, the gesture type, e.g. hand configuration, can be recognized to provide a set of discrete commands. While a computational system may only be reliable in recognition of a finite set of gesture types resulting in a finite gesture dictionary, concatenating basic gestures into patterns can broaden up the set of possible commands.

Sustained gestures always have a finite duration, which can be used as one computational parameter in a system. Pitched sustained gestures can convey information by pitch in various ways as discussed above. Short melodies can be mapped to discrete commands or functions, while gliding pitches can be used to tweak a continuous parameter in the system. In addition, the sound volume or its variation, and timbre can be tracked to enrich the information flow. Unpitched sustained gestures also are characterized by duration, and as shown by Harrison and Hudson [19], can be used to perform recognizable gestures based on temporal and timbral trajectories.

Looking at the parameter set, it can be argued that most of the sound-producing actions can be used for both discrete and continuous interactions. For impulsive actions, this usually requires performing an iterative gesture rather than one impulsive instance of the basic sound. For sustained actions to be used for discrete interactions, the solution is to consider them as objects rather than dynamic trajectories. This all boils down to mapping and selecting gestures with computable parameters that map well to the desired output parameters.

The gestures and their extractable parameters can also be studied in a multi-level hierarchy reflecting the complexity of the gestures and the extractable parameters at each level. This approach is presented in Table 2. At the lowest level, we have simple sonic gestures like individual hand claps, hums, and scratches. On the next level, here defined as the dynamic level, the body of the sounds can change its shape during the gesture, as for example in a whistle with a continuously changing pitch. Above these is the iterative level, i.e., all the iterative gestures. The highest level in this hierarchy is the compound level, which includes any combinations of the lower-level gestures, and can in theory provide an infinite group of potential gestures.

A possibility yet largely unexplored is to build interfaces around compound gestures and simultaneously occurring sonic gestures of different types. Compound sonic gestures could be gestural patterns, in which several gestures of different type follow each other (a finger snap followed by a whistle, for example). To facilitate richness of information at the interface, it would also be feasible to design interfaces where a sustained pitched gesture is used to control a continuous action, and an impulsive gesture that could overlap in the stream with the sustained gesture to indicate a discrete command. This approach would allow gestural multi-tasking.

Considering different sonic gestures, it is clear that they have different constraints in how they can be physically produced. Hand clapping or table tapping is easier to perform with fast tempos than finger snapping, for example. The same applies for sustained gestures, where for example the natural pitch range for humming is limited and even varies between different people, as shown by Sporka in [25], who found that the average comfortable pitch range is 12.7 semitones. Personal adjustability or system adaptability in fine-tuning the mapping can be useful in sonic interfaces.

Level of complexity	Impulsive	Sustained
Compound	Any of the below and their combinations	
Iterative	Tempo, temporal deviation, acceleration/deceleration, patterns **Example SGs:** Walking with a constant tempo, clapping a pattern of different hand clap types	Tempo, temporal deviation, acceleration/deceleration, patterns, melodies **Example SGs:** Scratching the table in a repetitive motion, humming a melody
Dynamic		Trajectories of changing pitch, timbre, volume, etc. **Example SGs:** Whistling with a rising pitch, scratching the table in an arc motion
Basic	Type, volume, timbre, direction **Example SGs:** Hand clap, finger snap	Type, pitch, duration, volume, timbre, direction **Example SGs:** Whistling with a constant pitch

Table 2. A multi-level presentation of extractable informational parameters from impulsive and sustained sonic gestures (SGs). The levels indicate the complexity of the gestures (and the required processing algorithms). On the basic level, we find simple gestures like individual claps, snaps, and hums. The dynamic level does not exist for impulsive gestures in this case, as their body cannot be dynamically varied after the sound has been generated. The compound level is a placeholder for arbitrary combinations of the lower-level gestures.

There are also computational constraints to consider. For example, it is in general not feasible to implement a real-time algorithm reliably differentiating between an arbitrary number of impulsive gesture types. Therefore, the interface designer needs to select an optimal set of gestures for the task. If needed, the gesture typology can be extended with compound gestures or gesture patterns.

An interesting prospect in using sonic gestures lies in sound-based positioning of the sound-performing human. With an array of microphones, it is possible to detect the direction, from which the sound arrives. While this was already demonstrated by two microphones in [11] for a single user, it is possible for example to separate the sound streams of different users with positioning information from a larger array.

5. DISCUSSION

It is without question that sonic gestures can bear lots of information useful in human-computer interaction. This is not to say that sound should or could always be used as the only input modality in interactive applications, but rather that the interaction could be enriched by more broadly acknowledging the use of sonic gestures as one input modality in the sensory fusion. Also, as sonic gestures are by nature typically embodied actions, they have potential in creating ”natural” interactions. This, however, is dependent on finding suitable interaction primitives that result in aligned multisensory perception, as discussed in [14] in the context of continuous sonic interaction.

Designing interfaces around sonic gestures can be seen as closely related to the ”traditional” design of sonic interactions. Indeed, for example basic design has been proved to be a usable tool for SID in designing sonic feedback [26, 14], and it can be hypothesized that similar techniques could be used to design interfaces with sonic input.

This study does not discuss the computational methods for information acquisition from sonic gestures. In general, it can be stated that and algorithm for detecting sonic gestures needs to be specialized into certain gesture types, for example classifying different impulsive gestures or tracking the pitch of humming. However, several algorithms suitable for sonic gestural interfaces are already available from different fields of study. For example in the field of music information retrieval, more and more focus has recently been put to implementing real-time algorithms for sound recognition, tempo and beat tracking, pitch tracking, etc. These tools can be efficient also in the acquisition of information from inherently non-musical sonic gestures, as exemplified for example in [22].

One important notion in discussing the utility of sonic gestures is to acknowledge their social acceptability. As it is generally understood that noise pollution is nowadays everywhere especially in urban life, do we want to add to the chaotic soundscape people interacting with their devices by clapping their hands or hissing through their teeth? This question is interesting and challenging for interface designers, who need to take into account the potential contexts where their designs are applied, and that a cultural change to facilitate the use of new interaction schemes ubiquitously takes time. It may be that sonic gesture interfaces have most applications in private conditions, or in social interaction applications where the users occupy the same space. Utilizing less intrusive sonic gestures and placing the microphone close to the sound production may provide a solution.

6. CONCLUSIONS

This study has demonstrated that sonic gestures, as sound-producing and information-bearing actions of a human, may be used in many ways to realize new kinds of interfaces for human-computer interaction, and that they are able to convey a very rich set of information. While some of the reviewed applications are musical, the gestures themselves are inherently often non-musical, at least until the context of interaction is introduced. Different sounds bear different kinds of information, which can be mapped in a desired way to the system output. While similar studies are known in the field of gestural interfaces, for sonic gestures a solid body of work demonstrating where different sonic gestures may be useful has not been previously presented.

An important aspect to consider when designing interfaces around sonic gestures is to aim for maximally natural and intuitive interaction. While it is entirely possible to derive continuous parameters from iterative gestures, for example, it does not mean that the mapping to any continuous output parameter is meaningful. To derive more comprehensive guidelines for facilitating the use of sonic gestures in the fields of HCI and SMC, one potential approach could be design patterns [27], capturing the use of sonic gestures in context to highlight good use scenarios, available computational techniques, and gestural relations. However, to date the body of work on sonic gestures for system input is not broad enough for deriving a comprehensive set of patterns.

An important prospect for future is to combine the presented gesture and parameter taxonomy with the physiological limitations of the production of different sonic gestures, e.g., what is the natural range of tempos for clapping and the natural pitch ranges in humming and whistling. This would underline the dynamic range of each gesture type and provide more guidelines for interface designers.

Acknowledgments

This work has been supported by the Graduate School at Aalto University School of Electrical Engineering, the Academy of Finland (project SCHEMA-SID), and the Finnish Foundation for Technology Promotion. The author would like to thank Dr. Cumhur Erkut for his support.

7. REFERENCES

[1] C. Cadoz and M. Wanderley, “Gesture-music,” in *Trends in Gestural Control of Music*. Paris, France: IRCAM - Centre Pompidou, 2000, pp. 71–93.

[2] E. Miranda and M. Wanderley, *New digital musical instruments: control and interaction beyond the keyboard*. Madison: AR Editions, Inc., 2006.

[3] A. Jensenius, *Action-sound: Developing methods and tools to study music-related body movement*. Faculty of Humanities, University of Oslo, 2008.

[4] D. Van Nort, “Instrumental Listening: sonic gesture as design principle,” *Organised Sound*, vol. 14, no. 02, pp. 177–187, 2009.

[5] R. Godøy and M. Leman, *Musical gestures: Sound, movement, and meaning*. New York: Taylor & Francis, 2009.

[6] M. Chion, *Le Son*. Paris: Editions Nathan, 1998.

[7] C. Cadoz, "Instrumental Gesture and Musical Composition," in *Proc. Intl. Computer Music Conf.*, Cologne, Germany, 1988, pp. 1–12.

[8] I. Ekman and M. Rinott, "Using vocal sketching for designing sonic interactions," in *Proc. 8th ACM Conf. Designing Interactive Systems*, Aarhus, Denmark, 2010, pp. 123–131.

[9] A. Dessein and G. Lemaitre, "Free classification of vocal imitations of everyday sounds," in *Proc. Sound and Music Computing Conf.*, Porto, Portugal, 2009.

[10] A. Sporka, "Pitch in non-verbal vocal input," *ACM SIGACCESS Accessibility and Computing*, no. 94, pp. 9–16, 2009.

[11] S. Vesa and T. Lokki, "An eyes-free user interface controlled by finger snaps," in *Proc. 8th Intl. Conf. Digital Audio Effects (DAFx)*, Madrid, Spain, 2005, pp. 262–265.

[12] A. Jylhä and C. Erkut, "A hand clap interface for sonic interaction with the computer," in *Proc. Conf. Human Factors in Computing Systems (CHI)*, Boston, MA, USA, 2009, pp. 3175–3180, presented in interactivity.

[13] A. Jylhä, I. Ekman, C. Erkut, and K. Tahiroğlu, "Design and Evaluation of Rhythmic Interaction with an Interactive Tutoring System," *Computer Music J.*, vol. 35, no. 2, pp. 36–48, 2011.

[14] D. Rocchesso, P. Polotti, and S. Delle Monache, "Designing continuous sonic interaction," *Intl. J. Design*, vol. 3, no. 3, pp. 13–25, 2009.

[15] J. Bilmes, J. Malkin, X. Li, S. Harada, K. Kilanski, K. Kirchhoffi, R. Wright, A. Subramanya, J. Landay, P. Dowden *et al.*, "The vocal joystick," in *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, Toulouse, France, 2006, pp. I–625–I–628.

[16] S. Chanjaradwichai, P. Punyabukkana, and A. Suchato, "Design and evaluation of a non-verbal voice-controlled cursor for point-and-click tasks," in *Proc. 4th Intl. Conv. Rehabilitation Engineering & Assistive Technology*, Las Vegas, NV, USA, 2010, pp. 48:1–48:4.

[17] P. Hämäläinen, "Novel applications of real-time audiovisual signal processing technology for art and sports education and entertainment," Ph.D. dissertation, Helsinki University of Technology, 2007.

[18] A. Hazan, "Performing Expressive Rhythms with BillaBoop Voice-Driven Drum Generator," in *Proc. 7th Intl. Conf. Digital Audio Effects (DAFx)*, Naples, Italy, 2004.

[19] C. Harrison and S. E. Hudson, "Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces," in *Proc. 21st annual ACM Symp. on User Interface Software and Technology*, ser. UIST '08, 2008, pp. 205–208.

[20] G. Wang, "Designing smules iphone ocarina," in *Proc. Intl. Conf. New Interfaces for Musical Expression*, Pittsburgh, PA, USA, 2009.

[21] A. Jylhä and C. Erkut, "Inferring the hand configuration from hand clapping sounds," in *Proc. 11th Intl. Conf. Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008, pp. 300–304. [Online]. Available: http://www.acoustics.hut.fi/dafx08/papers/dafx08_52.pdf

[22] U. Şimşekli, A. Jylhä, C. Erkut, and A. Cemgil, "Real-Time Recognition of Percussive Sounds by a Model-Based Method," *EURASIP J. Advances in Signal Processing*, 2011, special Issue on Musical Applications of Real-Time Signal Processing.

[23] W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[24] J. A. Ballas, "Self-produced sound: tightly binding haptics and audio," in *Proc. 2nd Intl. Conf. Haptic and Audio Interaction Design*. Berlin / Heidelberg: Springer-Verlag, 2007, pp. 1–8.

[25] A. Sporka, "Non-speech Sounds for User Interface Control," Ph.D. dissertation, Faculty of Electrical Engineering, Czech Technical University, 2008.

[26] K. Franinovic and Y. Visell, "Strategies for sonic interaction design: from context to basic design," in *Proc. 14th Intl. Conf. Auditory Display*, Paris, France, 2008.

[27] J. Borchers, "A Pattern Approach to Interaction Design," *AI & Society Journal of Human-Centred Systems and Machine Intelligence*, vol. 15, no. 4, pp. 359–376, 2001.