

Vocal Telekinesis; Physical Control of Inanimate Objects with Minimal Paralinguistic Voice Input

Sama'a Al Hashimi
Lansdown Centre for Electronic Arts
Middlesex University
Hertfordshire, United Kingdom
Samaa.alhashimi@gmail.com

Gordon Davies
Lansdown Centre for Electronic Arts
Middlesex University
Hertfordshire, United Kingdom
mail@gordondavies.com

ABSTRACT

Growing awareness of the possible over-dominance of the visual modality in the field of interactive media and of the existence of untapped dimensions of sound has led many developers to embark upon the development of sound-based projects. Many of these projects restrict themselves to delivering and fostering audio-visual and audio-physical art works. In this paper, we advocate the progression of audio-visual applications into voice-visual performances, and the evolution of audio-physical applications into voice-physical installations. We suggest new forms of voice-physical artwork aimed at the use of paralinguistic vocalizations to physically control real inanimate objects.

Categories and Subject Descriptors

H.5.5 Sound and Music Computing (e.g., HCI): Miscellaneous.

General Terms

Performance, Design, Experimentation

Keywords

Paralanguage, vocal input, vocal telekinesis, voice-physical

1. INTRODUCTION

Motion can cause sound, but can sound cause motion? If not, then can sound be programmed to cause motion?

This question has resonated in our minds every time we watched the table vibrate below the stereo's speakers while we tested our sound-based projects. The sound output from those speakers didn't only move the surface of the table but also directed our thoughts towards a way to use sound to physically move inanimate objects in real life.

Most people who have conducted the 'sound-vibrations experiment' with rice at school know that sound does actually cause motion; it causes objects to vibrate at the frequency of the waveform. Each object has a set of natural frequencies at which it may vibrate. While the stereo is playing loud music, the air around the stereo vibrates. If one of the sound frequencies generated from the stereo matches the natural frequency of a nearby object, resonance will occur. It is also known that some sopranos are able to shatter glass

with their voices. When the singer's voice is loud enough and when its pitch matches the resonance frequency of the glass, the glass may shatter. Another phenomenon that can be caused by resonance is the collapse of a bridge as a result of soldiers' rhythmic marching.

Within the field of interactive media, sound can also be programmed to "cause" any kind of movement. Nonetheless, most of the attempts to explore correlates of sound have been made to map acoustic characteristics to visual screen-based parameters. As a result, many sound-related installations are audiovisual applications while very few, in comparison, are audio-physical, and hardly any are voice-physical. In many of the few audio-physical applications which exist today, physical movement is what controls sound and not the opposite. Focusing on sound output, or even sound input, as a primary mean of interacting with an installation may allow the use of other modes of interaction simultaneously. It facilitates the avoidance of using the traditional mouse and keyboard. Using the acoustic mode as the main mode of interaction may reduce visual interaction with the screen, but it may also facilitate and open up possibilities for visual interaction with objects other than, and away from, the screen. William Gaver explains this in terms of attention directionality:-

"Visual objects exist in space but over time, while sound exists in time, but over space[...]One does not have to face a source of sound to listen to it. This implies that sound can convey information to users despite their orientation, while visual information depends on users' directed attention" [2].

Many audiovisual works, however, seem to neglect the fact that sound has overlooked dimensions that allow for its use as an input as well as an output. One neglected dimension is the paralinguistic dimension which includes voice characteristics (pitch, volume, timbre, etc.), emotional vocalizations (laughing, crying, screaming, etc), and vocal segregates (fillers like *ahh* and *mmm*, pauses, and other hesitation phenomena). Using this dimension of voice in an interactive application does not require access to the verbal content. This may expand the cross-cultural scope of the interactive work and allow its use by any users regardless of their language. The use of paralinguistic vocal control creates a real-time causal relationship between the acoustic input and the visual or physical, or haptic output, and may therefore facilitate continuity and direct engagement. It may also allow an installation to be used for therapeutic and vocal training purposes.

Through reversing the role of sound in an installation by making it a controlling factor rather than a controllable element, we hope to introduce the terms *voice-visual* and *voice-physical* to the field of interactive media and hence to enrich the repertoire of interaction.

In this paper, we will use the term *voice-visual* to refer to the use of voice as an input and the use of screen-based visuals as an output. The term *voice-physical* will be used to refer to the use of voice as an input and non-screen-based physical events as an output.

We started exploring the use of paralinguistic voice as an input to by developing a number of *voice-visual* games. One of these games is *Sing Pong* which is a voice-controlled version of 'Pong'. Whereas in the traditional Pong the arrow keys in a keyboard are used to move the paddles, *Sing Pong* allows players to move the paddles using their voices and shadows. The paddle's height is mapped to volume, while its position is mapped to the position of the player's shadow on a projected screen. Players' interaction with *Sing Pong* in an exhibition in London motivated us to further explore possible novel uses of paralinguistic voice. We soon realized that moving the players away from the monitor was not the only advantage that vocal input allows, as it also allows moving the output out of the monitor.

2. VOCAL TELEKINESIS

For our purposes, we define *Vocal Telekinesis* as the physical control of inanimate objects via simple paralinguistic vocal input. We aim to explore a variety of novel voice-physical mappings which will extend beyond the graphical output to include physical feedback such as changes in the size, temperature, brightness, speed, direction, and height of real objects. Our first implementation of *Vocal Telekinesis* was in *sssSnake*.

sssSnake is a two-player *voice-physical* version of the classic 'Snake' game (see Figure 1). It consists of an installation table on which a real coin is placed and a virtual snake is projected. One player controls the snake by uttering "SSS", while the other player moves the coin away from the snake by uttering "AAHHH". The position of a player round the table determines the direction of the coin's or snake's path. The snake moves towards the player uttering "SSS", and the coin moves away from the player uttering "AHH". Through prompting players to run round the playing area, the game encourages physical movement as well as vocal activity. The "SSS" and "AHH" voices are not differentiated through speech recognition but rather through the detection of frequency range differences between the high-pitched "SSS" and the low-pitched "AHH".

We programmed *sssSnake* in Lingo/Macromedia Director, using the asFFT Xtra (external software module) which employs the Fast Fourier Transform (FFT) algorithm [3]. The movement of the coin involved using the Hewlett-Packard Graphics Language (HPGL) to program a plotter to move its head in response to the "AAHH" voice. A magnet was attached to the plotter head, and the plotter was hidden below the surface of the installation table.

3. CONCLUSION AND FUTURE WORK

Many multimedia applications are used silently, and any vocalizations are usually made only while interacting with speech recognition software, during a Voice over Internet Protocol (VOIP)

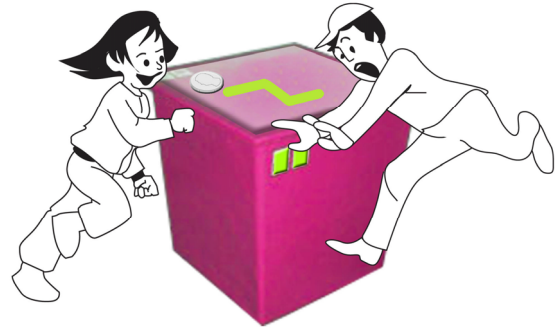


Figure 1. *sssSnake*: a voice-physical version of the classic 'Snake' game (2005)

audio conversation, or while cursing a crashed computer. Most of these vocalizations consist of speech and whenever non-speech sound is digitally involved, it is almost always generated by the computer in the form of beeps and similar sounds. We believe that voice is still partly mute in the field of interactive media. The use of its verbal dimension mainly as a communication tool overshadows many other possible uses. We hope that our *voice-visual* and *voice-physical* installations will blur the boundaries between input and output by enabling the voice to be both an input to the installation and an output to the audience. One of our goals is to use the various emerging non-conventional input mechanisms to exploit the potential of the body to be a rich source of input; rich enough to be an output too.

Moreover, most causality experiments to date have concerned a physical object causing another physical object to move, or a physical event causing sound (ex. Michotte's experiments in [1]). As no causality studies appear to exist in the voice-physical area addressed by our research, our future work will involve designing a usability test of the perception of causality in voice-controlled applications, with particular emphasis on the issue of latency.

We think that using voice as a physical tool pushes some of the boundaries of human-computer interaction and allows interaction with forms of output beyond conventional screen-based two-dimensional and three-dimensional visuals. It helps the interface extend beyond the monitor and become an integral part of the physical world.

4. REFERENCES

- [1] Cavazza, M., Lugin, J., Crooks, S., Nandi, A., Palmer, M., and Le Renard, M. Causality and virtual reality art. In *Proceedings of the 5th Conference on Creativity & Cognition* (London, United Kingdom, April 12 - 15). C&C '05. ACM Press, New York, NY, 4-12, 2005.
- [2] Gaver, W. The Sonic Finder: An Interface that Uses Auditory Icons. In *Human-Computer Interaction*, 4 (1) p. 67-94, 1989.
- [3] Schmitt, A. asFFT Xtra. <http://www.as-ci.net/asFFTExtra>, 2003.