THE "CALL CENTRE" INSTALLATION

Ian Gibson[1]
Faculty of Information and Technology
Leeds Metropolitan University

Alistair Edwards and Chris Newell
Department of Computer Science
The University of York

Stuart Andrews
School of Arts and New Media
Scarborough Campus
University of Hull

ABSTRACT

This paper describes an installation entitled "Call Centre". Users interact with it using a modified telephone and a number of hardware sensors. The installation provides the user with experiences of isolation and communication. The title "Call Centre" suggests a member of the public trying to become aware and navigate through different channels of communication, with varying degrees of success. The sensor input software system is written by Ian Gibson using MAX MSP. The speech synthesis system uses a new system S2:P3 created by Chris Newell based on his STEVS prototype. Stuart Andrews is writing theatrical scripts for the voices to perform. The main sound output is created using 3 audio channels. Each audio stream (speech, breaths and ambience) is directed to a different channel and then mixed. The. S2:P3 system uses an implementation of SSML (Synthetic Speech Mark-up Language) found in Microsoft Speech Application Programming Interface (SAPI) [4]

Keywords – call centre, installation, SAPI, SSML

1. INTRODUCTION

The "Call Centre" project includes three areas of research. The first area is that of using hardware sensors (such as proximity detectors, floor pressure pads and handheld motion detectors) to create audio installations. The second is area is that of speech synthesis, more specifically examining the use of pauses and their effect in performance. The third area is that of location in performance spaces to create virtual environments (e.g. suggesting nostalgia, isolation etc).

The paper describes the sensor system, the phone hardware, the software and future developments.

2. SENSOR AND PHONE HARDWARE

This section describes the hardware used for "Call Centre". The hardware is applied in 2 areas: to generate an ambient soundscape and to provide the main scripted audio for interaction. An overview of the system is given in Figure 1.

2.1. Hardware Sensors

The hardware used is based around the MIDIcreator System [1]. MIDIcreator enables a musician to trigger monophonic and polyphonic note sequences via MIDI. The main unit has sixteen inputs into which a wide variety of sensors may be connected (e.g. motion detectors, proximity sensors, floor pressure pads, percussive wood blocks etc). Each sensor may be generate either a range of values or an on/off signal.

MIDIcreator is designed for both able-bodied and disabled users. The variety of sensors available allows even the smallest movement to be detected. The system is flexible enough to allow users to play individually or in groups. It encourages users to explore methods of making music other than with a traditional keyboard. It can be used live (with software such as MAX MSP) or users can record output to a sequencer in order to compose in non real-time.

MIDIcreator is configured via software on the PC. The resulting patches are stored to memory card which can be plugged into the unit.

2.2. The Modified Telephone

The shell of telephone is a standard GPO push-button telephone from the 1970's. Additional electronics have been installed to send MIDI note-on/off messages via the telephone keypad. A mini loudspeaker has also been added. Although not implemented in the current software the original microphone has been replaced with an ex-military noise cancelling microphone to facilitate speech recognition in later software revisions (see future developments). The retro/lo-fi sound quality has been retained deliberately.

3. THE MAX MSP SYSTEM

The MAX MSP system [3] is used to generate ambient soundscapes and to provide audio input into the modified telephone. One part of the system monitors user activity. This system uses an ultrasonic proximity detector to sense the presence (or absence of) someone near to the installation. If there is no activity then, after

[1] Email: I.Gibson@leedsmet.ac.uk

an elapsed time, the system produces ambient sounds which are designed to encourage interaction. If the system senses activity (via the proximity sensor) then it then it fades any sound being emitted from the external speakers and starts to output to 2 channels routed to the inside of the telephone and to the telephone handset respectively. The handset speaker then outputs sounds generated by the MAX MSP system and by the S2P3 system (mentioned later).

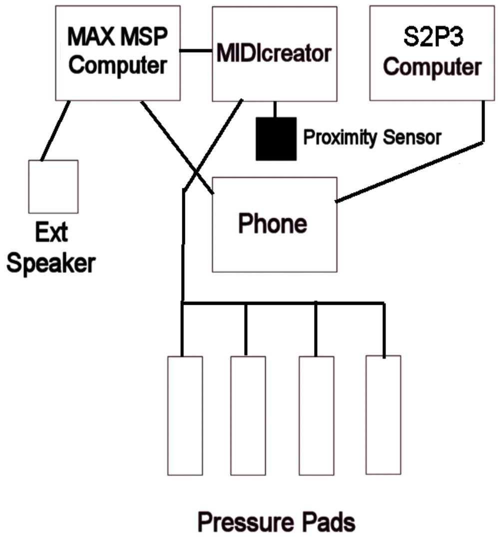

Figure 1. "Call Centre" Hardware

The MAX MSP system has four floor-based pressure sensors which allow the user to trigger ambient sounds; the audio is routed to the headset. The sounds complement those being produced by the S2:P3 system.

4. S2P3

Synthetic Speech: Pauses, Place and Persona (S2:P3) is designed as a quick enhancement to any Text to Speech Synthesis (TTS) system.

Research in speech synthesis has tended to concentrate on developing realistic voices, very often by using samples of real human speech and cleverly stitching them together (concatenative synthesis). This has proved data intensive and expensive. Recent experiments indicate that users accept a degree of unnaturalness in some aspects of a computer generated voice but are surprisingly fussy about other aspects [5]. In other words a cheap robotic voice that performs in a certain way can be more acceptable to users than an expensive, realistic voice that sounds dull or unappealing.

S2:P3 has been developed from research into human performers, particularly classical actors, singers and comedians [6]. In these genres pauses can be used to persuade the audience that the performer is inventing the words afresh rather than repeating the words in the script. S2P3 takes this a stage further adding breaths, ambience and non-grammatical (random) pauses to the synthetic speech stream. No other adjustments are made although SSML could be used to provide broad brushstroke changes to the prosody. The result is not intended to be more realistic but more spontaneous, thoughtful or with added 'liveness.'

5. FUTURE DEVELOPMENTS

Newell and Gibson are planning a live performance of an enhanced version of the system. Interaction is to be controlled by speech recognition alongside the existing midi controls and MAX MSP. The new work entitled 'Microsoft Mary's Comic Potential' is based on the play 'Comic Potential' by Alan Ayckbourn [2]. The stage performance will be cooperatively delivered by a live human mimic and the S2:P3 system.

6. ACKNOWLEDGEMENTS

The authors would like to thank the University of Hull for their financial support of this project.

7. REFERENCES

[1] Abbotson, M & R, Kirk, P.R., Hunt, A.D. & Cleaton, A. Computer Music in the service of Music Therapy: The MIDIGRID and MIDICREATOR systems, Medical Engineering Physics, Vol 16, May 1994, pp 253.

[2] Ayckbourn, A. (2002). Comic potential. New York; London, Samuel French.

[3] Cycling 74: http://www.cycling74.com/

[4] Microsoft.(2005). SAPI. http://www.microsoft.com/downloads/details.aspx?FamilyId=5E86EC97-40A7-453F-B0EE-6583171B4530&displaylang=en

[5] Nass, C. I. and S. Brave (2005). Wired for speech: how voice activates and advances the human-computer relationship. Cambridge, Mass., MIT Press.

[6] Newell, C. and A. Edwards (2005). "Lively voice: a new model for speaking synthetic characters." International Journal for Language Data Processing 28.2/2004: 133-151.