SQEAK: Using your Mobile Phone as a Gesture Sensor

Dennis Majoe, Stefan Müller Arisona, Art Clay
Computer Systems Institute
ETH Zurich
[dennis.majoe I stefan.mueller I art.clay]@inf.ethz.ch

Simon Schubiger
Swisscom Innovations
simon.schubiger@swisscom.com

ABSTRACT
This research looks at one approach to providing all mobile phone users with a simple low cost real time user interface allowing them to control highly interactive public space applications involving one user or a large number of simultaneous users. In order to sense accurately the real time hand movement gestures of mobile phone users the method uses miniature accelerometers that send the orientation signals over the network's audio channel to a central computer for signal processing and application delivery. This affords that there is minimal delay, minimal connection protocol incompatibility and minimal mobile phone type or version discrimination. Without the need for mass user compliance, large numbers of users could begin to control public space cultural and entertainment applications using simple gesture movements..

Keywords
Computer interaction, gesture recognition, ubiquitous computing, public entertainment, public information, mobile phone.

1. INTRODUCTION
Increasingly, pervasive computing is allowing mobile users to do more with their mobile devices. A number of Mobile Social Software based systems are emerging [1] that allow users to come into contact based on their proximity however these public social applications do not open up possibilities for members of the public interacting with outdoor ubiquitous computing. For example mobile users could be directly interacting with ubiquitous computing devices such as train or plane timetable plasma screens, electronic advertisement displays and shop window systems. In the same way, large assemblies of people in concerts, sports events and shopping malls, cannot yet express themselves as a social collective, interacting with a shared expression. Highly interactive applications that involve and inspire users en masse are not at all common. This shortfall may be due to a number of factors, in particular the lack of an intuitive interactive device in the mobile phone and the fact that mobile technology has not sufficiently converged to provide the majority of users with interoperable interaction devices, allowing public interaction applications to flourish.

2. MOBILE PHONE LIMITATIONS
Interaction with a mobile phone in a public space should be as natural as possible. Using the keypad may be practical but is rather mundane. Using natural speech recognition is difficult enough let alone over significant background noise. The camera has been used in several research projects to provide a pointing device capability by processing changes in the camera picture recorded during the phone's movement but this needs good lighting. [2] Recently accelerometers have been included in a mobile phone by Samsung and research performed on the use of such a device [3] showed that gesture recognition of user movement can be used to drive the mobile phone's applications and local games. However in this case, one must send this data over the network and a number of problems would be encountered [4]. First of all phones would have to adopt a standard data transmission protocol and a standard packet description for the sensor data. Each phone would have to adopt this very same protocol otherwise one application would need to support a plethora of protocols. Secondly as digital data transmission has a lower priority to voice data, there is no guarantee that sensor data will arrive in true real time. Finally few younger users would elect for anything more than voice and text messaging, as part of the mobile phone service and this therefore would exclude many of the most likely early adopters.

3. THE SQEAK
What is required is an extremely simple way of converting any normal mobile phone into a gesture motion sensor device compatible for use on any network. The concept behind the SQEAK was to develop an ultra low cost device, the size of a shirt button, that would stick onto any mobile phone, just next to or on top of the microphone input hole. The button would contain accelerometer sensors, electronics and battery and would generate an audio tone that would be sent down the audio channel of the phone. This means any phone can be used and motion data is sent at high voice priority level. The tone would contain one, two or more superimposed audio signals at different frequencies. Each signal would be frequency modulated about a central frequency as a function of the accelerometer signal amplitude in any axis. In this way a single audio signal would carry in real time the accelerometer sensor data to a central application server for decoding allowing any phone to be used on any network as a general purpose gesture sensor. The audio signals would need to be such that they passed through the voice signal coding and decoding processes of the mobile phone system and still be reliably used for gesture recognition in real time.

Fig. 1 Proof of concept prototype connected to a mobile phone.

3.1 Prototype

The first working prototype was based on a very low cost voltage controlled oscillator (VCO) with plastic, low cost mechanical accelerometer sensors. The aim was to meet a sub 25 US cents manufactured cost. In the second testing prototype, shown in Fig. 1, a PIC microcontroller was used to read the two voltage outputs of an Analog Devices 2 axis accelerometer. The firmware was designed to generate two digital outputs at two centre frequencies. The outputs would then be modulated as a function of the accelerometer output voltages and used to drive a small microphone.

3.2 Signals Over the Network

The following shows the frequency spectra of signals sent over the network for the SQEAK at different angles of inclination. As one can see the digital signals are reliably converted to typical sinusoidal single band frequency modulated (FM) signals, which encode the inclination measurement. This occurs despite the network's human voice adaptation idiosyncrasies.

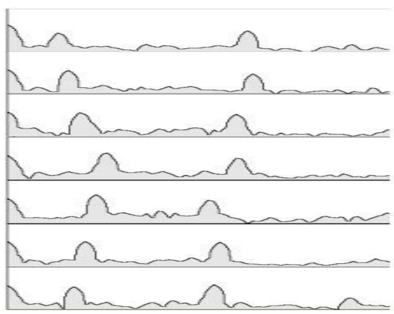
0 KHz  2 KHz

Fig. 2: Spectra of two-axis FM signal at different angles.

3.3 Signal processing

In order to process the audio signals, processing software was written based on the versatile audio processing libraries available in Max/MSP. A combination of bandpass, low and high pass filters are used to separate the two different axis signals before routing to two pitch detectors.

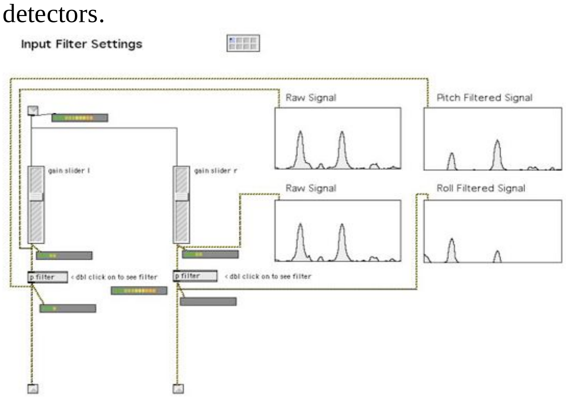


Fig 3: Filters enhance each of the two centre frequencies prior to dual pitch detection

The pitch detector used is an object called "fiddle" written by Miller S. Puckette, [5] one of the original developers of Max. It is normally used to detect pitch classes in musical instruments. The fiddle object, although it was not designed for this application, has been useful because it assumes that there are more than just one principle component in the frequency spectrum. Therefore it returns the number of peaks, the frequency of each and the strongest component first.

3.4 A Simple Public Demo

In order to demonstrate how the SQEAK might function as a controller for an outdoor public display, a simple art based application was developed. The concept behind the application was to use the art concepts of Josef Albers. This entailed creating an Albers-like painting with vector graphic having particular formal relationship based on contrasts of form, color, and size. Basically, changes to parameters for color and form are made on a dynamic level based on the audio output from the SQEAK device. Using the pitch and roll axis of the SQEAK, the user can interact with the graphic to explore the relationships between the individual elements of the graphic in real time.

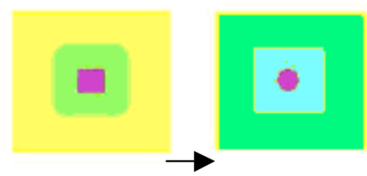


Fig. 4: The 2 axis signal controls the color and geometry

4. ACKNOWLEDGMENTS

Our thanks go to Swisscom for their cooperation in this proof of concept work.

5. REFERENCES

[1] http://www.dodgeball.com/.

[2] Rafael Ballagas, Michael Rohs, Jennifer G. Sheridan, "Mobile Phones as Pointing Devices", Pervasive Mobile Interaction Devices (PERMID 2005), Workshop at the Pervasive 2005, Munich, Germany

[3] Sung-Jung Cho, Jong Koo Oh, Won-Chul Bang, Wook Chang, Eunseok Choi, Yang Jing, Joonkee Cho, Dong Yoon Kim, "Magic wand: a hand-drawn gesture input device in 3-D space with inertial sensors", Samsung Adv. Inst. of Technol., Suwon, South Korea. Ninth International Workshop on Publication Date: 26-29 Oct. 2004

[4] Nilantha Katugampala, Khaldoon Al-Naimi, Stephane Villette, and Ahmet Kondoz, "Real Time End to End Secure Voice Communications over GSM Voice Channel", 13th European Signal Processing Conference, Antalya, Turkey, September 2005

[5] Pukette, M. "Max at seventeen", Computer Music Journal, 26(4): pp 31–43., 2002