Shoogle: Excitatory Multimodal Interaction on Mobile Devices

John Williamson[1]
[1] Dept. Computing Science
University of Glasgow
Scotland
jhw@dcs.gla.ac.uk

Rod Murray-Smith[1,2]
[2]Hamilton Institute
NUI, Maynooth
Ireland
rod@dcs.gla.ac.uk

Stephen Hughes[1,3]
[3]SAMH Engineering Services
Dublin
Ireland
stephenahughes@gmail.com

ABSTRACT

Shoogle is a novel, intuitive interface for sensing data within a mobile device, such as presence and properties of text messages or remaining resources. It is based around active exploration: devices are shaken, revealing the contents rattling around "inside". Vibrotactile display and realistic impact sonification create a compelling system. Inertial sensing is used for completely eyes-free, single-handed interaction that is entirely natural. Prototypes are described running both on a PDA and on a mobile phone with a wireless sensor pack. Scenarios of use are explored where active sensing is more appropriate than the dominant alert paradigm.

Author Keywords

Audio, vibrotactile, accelerometer, multimodal, mobile

ACM Classification Keywords

H.5.2 User Interfaces: Haptic I/O; H.5.2 User Interfaces: Interaction Styles; H.5.2 User Interfaces: Auditory (non-speech) feedback

> *shoogle v.*
> *1. intr. To sway, move unsteadily, to rock, wobble, swing*
> — The Dictionary of the Scots Language

MOTIVATION

Users of mobile devices are continuously bombarded with alerts and notifications; waiting voice mail messages, incoming SMS messages, low battery alarms and so on. This disrupts the tasks they are otherwise engaged in and can often be socially unacceptable. We propose an alternative interaction style, where the user *excites* information from the device and then *negotiates* with the system, in a continuous, closed-loop interaction.

Model-based sonification was introduced by Hermann and Ritter [3]. They state: *"...why not sonify data spaces by taking the environmental sound production in our real world as a model. Nature has optimized our auditory senses to extract information from the auditory signal that is produced by our physical environment. Thus the idea is: build a virtual scenario from the data; define a kind of 'virtual physics' that permits vibrational reaction of its elements to external excitations; let the user interactively excite the system and listen."* This work applies this concept to the display of the internal state of mobile devices: sonifying and haptically rendering the contents of inboxes, the state of battery life, or remaining memory. Feedback is tightly coupled to the input. In contrast to static display approaches (such as vibration alerts), this *active perception* approach takes advantage of people's expectations about the evolution of dynamic systems. This avoids interrupting or disturbing the user unnecessarily and opens up the potential for richer, more informative feedback. Users know what motions they have applied and interpret the display in that specific context.

Impact perception is a task with which everyone is familiar; few people would have difficulty distinguishing a hollow barrel from a full one after tapping it. Because such information is communicated primarily through the auditory and haptic channels, a completely non-visual interaction can be constructed. Given that mobile devices are often used where visual attention is inconvenient, the use of purely non-visual cues is a major advantage over visually-dominated techniques. The Shoogle interface uses inertial sensing for natural motion sensing without any external moving parts; the user just shakes, tilts or wobbles the device to stimulate the auditory and vibrotactile feedback. This can either be an explicit action, or can occur as part of a user's background motion; walking, running, standing up or other everyday motions. The interaction transforms the system from a sizeless portal through which information flows to an embodied container, with physically meaningful characterstics. The following scenarios illustrate how a device augmented with these capabilities could be used in everyday tasks, while remaining eyes-free and without interrupting the user.

Scenario: Eyes-Free Message Box

The user reaches into his or her bag, and shakes the phone gently, without removing or looking at it. The contents of the SMS inbox are transformed into virtual "message balls". As the user shakes it, impacts are heard and felt as these balls bounce around. Hearing that one of them is distinctly metallic, and has a deep, heavy feeling impact, the user realises

CHI 2007, 28 April - 3 May, 2007, San Jose, California, USA.

that a long text message from a colleague has been received. The sensation of a number of lighter, glassy objects indicates several short messages from various friends. As messages arrive, they "drop" into the device, with appropriate sound and dynamics.

Scenario: "Keys" in a Pocket

The user carries the phone in a pocket while walking. Gait motion is sensed with the accelerometers. As messages arrive, objects begin jangling around, in a manner similar to loose change or keys. This subtly notifies the user of the presence of the messages – a comforting sense of presence rather than an intrusive alarm. This is an example of the "background interaction" proposed by Hinckley *et al.* [4]. The specific audio properties reveal elements of the content; for example having longer messages sound like the heavy iron keys of a jailor, and shorter ones like small coins.

Scenario: Liquid Battery Life

The user shakes the device to gain a sense of its "fullness". A liquid metaphor is used. When the battery is full, the sensation is like that of a full bucket of water sloshing around. As the battery drains, shaking the device sounds like a few droplets splashing, until finally all power evaporates. This is similar to the virtual maracas approach for resource sensing suggested by Fernström in [2].

Many other scenarios could be envisaged: shaking an iPod to get an overview of the genres in a playlist; exciting a contacts list, then sieving out potential interesting names; sensing upcoming events in a schedule, sonifying their priority; or sensing the files in memory, classified by size and type.

Selection

Although the prototypes are display-only, selection could be introduced via active selection techniques such as those presented in [14] and [13]. In such a configuration, the dynamics of each message or item of interest would have associated unique resonant modes which the user stimulates to gain more information. This could, for instance, involve directional resonances across particular planes of motion. Non-visual filtration and sorting of messages could be introduced in a similar manner; users pan or sieve the device for their particular "gold", using their intuitive understanding of the properties of the virtual objects to guide the process.

Background

Realistic synthesis of vibrotactile and audio sensations are key to building eyes-free interfaces. There has been a great deal of recent interest in physical models of contact sounds and associated vibration profiles. The model-driven approach is a fruitful design method for creating plausible and interpretable multimodal feedback without extensive *ad hoc* design. Yao and Hayward ([15]), for example, created a convincing sensation of a ball rolling down a hollow tube using an audio and vibrotactile display. A similar sonification of the physical motion of a ball along a beam is described in detail in [10]; subjects were able to percieve the motion ball from the sonification alone. A simple haptic "bouncing ball" on mobile devices was demonstrated by Linjama et al. [8], which used tap sensing to drive the motion of a ball; this, however did not incorporate realistic dynamics or auditory feedback. Granular approaches to realistic natural sound generation were explored in [9], where contact events sensed from a contact microphone above a bed of pebbles drove a sample-based granular synthesis engine. A wide variety of sonorities could be generated as a result of physical interactions with the pebbles. This granular approach is used as the synthesis engine in the Shoogle prototypes. Complete working prototypes of Shoogle have been implemented on both PDA and standard mobile phone platforms. There are three important components of Shoogle: the sensing hardware, to transform motion into measureable signals; the dynamics model, to simulate the motion of the virtual objects; and the display, which reveals the state of those objects.

Figure 1. The MESH expansion pack, with an iPaq 5550 PocketPC. This provides accelerometer, gyroscope and magnetometer readings, as well as vibrotactile display.

Figure 2. The wireless SHAKE sensor, shown with a 2 Euro piece for size comparison. This Bluetooth device comprises a complete inertial sensing platform.

INERTIAL SENSING

Reliable sensing of device motion is critical. This can be achieved by instrumenting the device with tri-axis accelerometers. Accelerometers have previously been widely used for tilting based interfaces (e.g. in [11] and [5]); in the present application, the linear acceleration component is more relevant than gravitational effects. The initial prototype of Shoogle was implemented on the iPaq 5550 device (see Figure 1), using the MESH ([6]) device for inertial sensing and on-board vibrotactile display. The MESH's vibrotactile transducer is a VBW32 loudspeaker-style device. This device allows for the display of high fidelity tactile sensations due to its large bandwidth and fast transient response. The iPaq's internal eccentric motor vibration unit is also used for vibrotactile display. This provides a lower frequency range of vibration sensations than achievable with the VBW32, like a woofer would in an audio system. The accelerometers are used for sensing, sampled at 100Hz, with a range of approximately $+/-$ 2g. Capacitive sensing is used for tap detection.

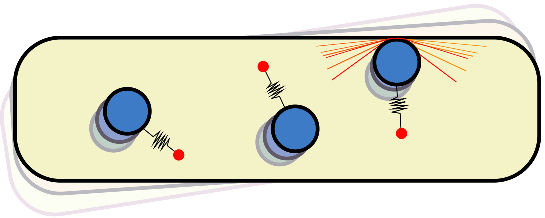

Figure 3. The simulated system. A number of balls, anchored via springs, bounce around within the virtual container. When they impact (as in the top right) sound and vibration are generated.

Shoogle on Mobile Phones – The SHAKE

A version of Shoogle has also been implemented on standard mobile phone hardware (Nokia 6660 and other modern Series 60 phones). This uses the custom-built Bluetooth SHAKE (Sensing Hardware Accessory for Kinesthetic Expression) inertial sensor pack for sensing. Mobile phones which already incorporate accelerometers (like the the Nokia 5500 and Samsung SCH-S310) would require no additional hardware. The SHAKE model SK6 is a small form factor wireless sensor-pack with integrated rechargeable battery, approximately the same size as a matchbox (see Figure 2). The device is a pre-production model and is now available to researchers on a commercial basis. It features tri-axis accelerometer, tri-axes angular rate sensor, tri-axis magnetometer, dual channel analog inputs, dual channel capacitive sensing and an internal vibrating motor. Communications are over a Bluetooth serial port profile. SHAKE includes a powerful DSP engine, allowing real time linear phase sample rate conversion. These capabilities allow rapid prototyping of inertial-sensing-based interfaces with real-world hardware. In the iPaq prototype, the arrival of messages is simulated; with the SHAKE-based phone prototype, the real SMS inbox can be directly accessed. The improved form-factor and access to true messaging and resource functionality makes completely realistic test scenarios possible.

DYNAMICS MODEL – MAKING THE BALLS BOUNCE

The Shoogle metaphor involves a number of spherical balls rattling around in a rectangular box whose physical dimensions appear to be the same as the device. The box is assumed to be two-dimensional, with no vertical component.

The simulated motion of the virtual objects is relatively simple. The measured accelerations are used directly (i.e. $\ddot{x}_q = a_q$ for each axis q) in an Euler integration model. Nonlinear frictional damping (a stiction model with different static and moving coefficients) is applied, which can be varied to simulate different "materials" inside the virtual box.

As each ball is generated (e.g. as a message arrives), it is anchored to a randomly-allocated position within the box by a Hooke-law spring (so $\ddot{x} = a_q + \frac{k(x-x_0)}{m}$, where x_0 is the anchor point). The spring coefficient k loosens or tightens the motion of the balls (see Figure 3).

Collision Detection

Feedback events are generated when the balls collide with the walls of the device "box". Inter-ball collisions are not tested for, as they have little effect but significant computational cost. Wall collisions are inelastic, transferring some kinetic energy to the wall, and the remainder to rebound. The rebound includes directional jitter – simulating a slightly rough surface – to reduce repetitive bouncing. Users can also tap the device (detected via capacitive sensing), at which the balls bounce upwards (the only case where the third dimension is considered). They then rain back down in the order of creation, giving a structured view of the contents. In the SMS message prototype, the mass of each ball is proportional to the length of the message. Longer messages result in heavier balls with appropriate dynamics. Other transformations could be introduced based upon the content of the message: the language used, the apparent "positivity" or "negativity" of the text, or the presence of keywords.

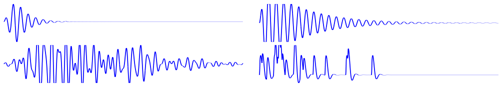

Figure 4. Four different vibrotactile waveform types. Left-to-right, top-to-bottom: standard "light ball" impact; heavy, ringing impact; liquid sloshing; gritty, particulate impact. All of these have energy concentrated around the 250Hz band.

AUDITORY AND VIBROTACTILE DISPLAY

The presentation of timely haptic responses greatly improves the sensation of a true object bouncing around within the device over an audio-only display. As Kuchenbecker *et al* [7] describe, event-based playback of high-frequency waveforms can greatly enhance the sensation of stiffness in force-feedback applications. In mobile scenarios, where kinaesthetic feedback is impractical, event-triggered vibration patterns can produce realistic impressions of contacts when the masses involved are sufficiently small.

The vibrotactile waveforms are enveloped sine waves, with center frequency at 250Hz (both the resonant frequency of the transducer, and the peak sensitivity of the skin receptors involved in vibrotactile perception). The waves have a very rapid linear attack, and an exponential decay. Variations on this basic theme have also been implemented. Sloshing-like vibrations are generated with waveforms having much longer attack portions, and slightly varying frequencies. Granular effects are created by summing a number of extremely short enveloped sine waves together, creating a dense, gritty texture. Heavier impacts are simulated with longer ringing portions and initially saturated output. Figure 4 gives an overview. Even when the device is empty (but actively "listening"), gentle vibration feedback is produced in response to shaking to indicate that the system is live.

To enhance the sensations, the internal eccentric motor of the iPaq is used in combination with the VBW32 transducer. Short, high-frequency vibrations are sent to the smaller vibrator, with heavy, slow impacts also being routed to the motor-driven actuator. The greater power of the motor-driven actuator results in a more "solid" feel, but limited control restricts output to simple events played in conjunction the high-frequency events (similar to the layering of sub-bass waveforms under recorded effects in film sound).

Audio Feedback

The impact of the balls on the virtual box produces sound related to the physics of the collisions. Although ideally these sounds would be generated by a physical model (such as the general contact sound engine given by van den Doel *et al* [12]), the limited computational power of many mobile devices – which lack efficient floating-point units – makes this currently impractical. Shoogle instead employs a sample-based technique. This gives a high degree of realism, butat the cost of significant effort and limited flexibility. A number of impact sounds (8–16) are pre-recorded for a particular impact "class" (wood on glass, for example). These slight variations are critical to avoid artificial sounding effects. On impact, a random sound from within this class is selected and mixed into the output stream. The audio output (which uses the FMOD library) mixes up to thirty-two simultaneous channels, to ensure that impacts do not cut off previous audio events in an unnatural manner.

Shoogle currently has eighteen impact types, including ping-pong balls hitting wood, candy rattling in jars, keys jangling and water sloshing in bottles. These provide a wide range of natural sounding and easily distinguishable timbres. Humans are exceedingly adept at inferring the physical properties of materials from the sound of their physical interaction, and the realistic nature of the generated sounds makes the nature of the impact immediately obvious.

Audio Transformations

The audio is transformed based on the properties of each particular collision. The intensity of the impact sound is proportional to the kinetic energy of the impact ($\frac{1}{2}mv^2$). The sample is pitch-shifted in proportion to the mass of the ball which impacted, so that large-mass interactions produce lower sounds with deeper resonances than smaller ones. Other transformations could be implemented, given sufficiently powerful hardware. Impulse response convolutions are widely used for high-quality reverberation; real-time convolution with the impact sounds could be used to simulate different container types with a high degree of realism.

Semantic Sonification

For the SMS scenario, the class of impact sound is linked to the content or the meta-data of the message represented, and thus the composition of messages in the inbox can be perceived as the device is manipulated. In the prototype, messages are tagged according to sender group (work, friends, family, unknown, etc.), where each group has a unique associated impact material. Messages could also be automatically classified with a language model (e.g. as in [1]), and the various styles or languages used could be mapped to different simulated materials.

CONCLUSIONS

The Shoogle prototypes illustrate how model-based interaction can be brought into practical mobile interfaces. The resulting interface is based around *active sensing*, letting the user drive the interaction. The result is a rich multimodal display that can be used without any visual attention whatsoever, taking advantage of user's familiarity with the dynamics of processes in the physical world to present information in a natural and non-irritating manner. The SHAKE sensor allows realistic inertial sensing prototypes to be rapidly implemented on mobile phones, with plausible form factors and all of the accumulated data a real phone carries. The ideas can be extended to include multimodal selection, extending the interaction from excitation to true negotiation. Although these prototypes only scratch the surface of the potential this interaction style offers, even in its current state the system is compelling. Of the several dozen people who have experimented with Shoogle, all have found it captivating and intuitive. Enormous scope exists to build sophisticated, model-driven interfaces upon this foundation, linking content and context to the physical properties of simulated systems, and the rich feedback which is a natural consequence of their behaviour.

Acknowledgements

The authors are grateful for support from: the IST Programme of the European Commission, via the *OpenInterface* project and PASCAL Network of Excellence, IST 2002-506778; IRCSET BRG SC/2003/271, Continuous Gestural Interaction with Mobile devices; HEA project *Body Space*; and SFI grant 00/PI.1/C067.

REFERENCES

1. P. Eslambolchilar and R. Murray-Smith. Model-based, multimodal interaction in document browsing. In *Multimodal Interaction and Related Machine Learning Algorithms*, 2006.
2. M. Fernström. Sound objects and human-computer interaction design. In D. Rocchesso and F. Fontana, editors, *The Sounding Object*, pages 45–59. Mondo Estremo Publishing, 2003.
3. T. Hermann and H. Ritter. Listen to your data: Model-based sonification for data analysis. In *Advances in intelligent computing and mulimedia systems.*, pages 189–194. IIASRC, 1999.
4. K. Hinckley, J. Pierce, E. Horvitz, and M. Sinclair. Foreground and background interaction with sensor-enhanced mobile devices. *ACM Trans. Comput.-Hum. Interact.*, 12(1):31–52, 2005.
5. K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz. Sensing techinques for mobile interaction. In *UIST'2000*, 2000.
6. S. Hughes, I. Oakley, and S. O'Modhrain. Mesh: Supporting mobile multi-modal interfaces. In *UIST 2004*. ACM, 2004.
7. K. J. Kuchenbecker, J. Fiene, and G. Niemeyer. Improving contact realism through event-based haptic feedback. *IEEE Transactions on Visualization and Computer Graphics*, 12(2):219–230, 2006.
8. J. Linjama, J. Hakkila, and S. Ronkainen. Gesture interfaces for mobile devices - minimalist approach for haptic interaction. In *CHI Workshop: Hands on Haptics: Exploring Non-Visual Visualisation Using the Sense of Touch*, 2005.
9. S. O'Modhrain and G. Essl. Pebblebox and crumblebag: Tactile interfaces for granular synthesis. In *NIME'04*, 2004.
10. M. Rath and D. Rocchesso. Continuous sonic feedback from a rolling ball. *IEEE MultiMedia*, 12(2):60–69, 2005.
11. J. Rekimoto. Tilting operations for small screen interfaces. In *ACM Symposium on User Interface Software and Technology*, pages 167–168, 1996.
12. K. van den Doel, P. G. Kry, and D. K. Pai. Foleyautomatic: physically-based sound effects for interactive simulation and animation. In *SIGGRAPH '01*, pages 537–544. ACM Press, 2001.
13. J. Williamson. *Continuous Uncertain Interaction*. PhD thesis, Dept. Computing Science, University of Glasgow, 2006.
14. J. Williamson and R. Murray-Smith. Pointing without a pointer. In *ACM SIG CHI*, pages 1407–1410. ACM, 2004.
15. H.-Y. Yao and V. Hayward. An experiment on length perception with a virtual rolling stone. In *Eurohaptics 06*, 2006.