Metaphor, Modality, and Context: Improved user interface design for audio and music applications

Andy Sarroff
Music Technology Department
New York University
andy.sarroff@nyu.edu

ABSTRACT

The sophistication of modern audio and music computing environments has placed new demands on informational sharing and interchange between computers and users. Multiple input/output modalities, better use of metaphor, and implementation of context awareness can smoothen and enhance a user's experience. This paper explores how these changes are being implemented in musical, auditory, and related fields. A model for interface design is suggested.

Keywords: Human-Computer Interaction (HCI), User Interface Design (UID), contextual awareness

1 INTRODUCTION

The audio and computing industries are offering new ways for interaction between computing environments and users. Technological advances have improved the depth and quality of information that can be conveyed to computers. Additionally, the informational throughput (and data rate) between users and computers has greatly increased. As a system, humans and computers are becoming more integrated and sophisticated in the management of information. Human-Computer Interaction (HCI) has emerged as one of the defining elements of interface design in the computing industry. As ubiquitous computing becomes imminent, the human-computer system requires further refinement. Computers are increasingly adapting to our language. However, there is still great room for progress in further facilitating the human-computer communication channels.

The success of future interface design is dependent on building transparency and fostering creativity. The creative use of technology necessitates adaptability even after an interface is placed in the hands of an operator. Over time, the usability of a device shifts as a person's relationship with that device changes in complex ways [1]. Successful designs anticipate an ever-changing association between computers and their human operators.

There are many ways in which interface designs can be engineered for the increasingly sophisticated computing resources available. New interfaces should allow for easy creativity, changing usability, and bidirectional intelligibility. Section 2 of this paper defines Human-Computer Interaction (HCI) and discusses HCI's relevance to good audio and musical interface design. Section 3 introduces metaphor as a design tool. In Section 4, modalities of input and output are explored, and implementations of multimodality are reviewed. Contextually aware methods are examined in Section 5. In Section 6 some recent, integrative interfaces for audio and music are canvassed. Finally, a model of interface design is synthesized from the HCI concepts reviewed in the paper.

2 HUMAN-COMPUTER INTERACTION

HCI is a human-centered approach toward computing, which aims to ease the cognitive load associated with interacting with computers. An unsophisticated example of HCI is an initiation of a web-based query in natural English syntax. A computer program is able to parse the meaningful content in a query from the grammatical connecting syntax that forms the query into a complete question. Although it is clearly useful for computers to understand natural grammatical English (or other languages), there are much more powerful channels of communication to examine. In [2], it is shown that it is insufficient for computers to merely understand what the content of a message is. For successful HCI, especially in ubiquitous computing environments, computers will need to interpret how a message is being passed on, and what the context of that message is. Computers, then, need to understand human *behavior*, as well as human *language*.

Behavior awareness is a complex task. A sliver of human expression can communicate a great deal of information. The quantization of vocal, facial and bodily gesture, along with a concept of context is no small undertaking. Complicating the matter is that there is no universal database of human expressions and gestures [3]. In fact, collecting an accurate database of such expressions presents significant problems of its own. Given that successful behavior awareness is possible, however, a system is left with a multitude of input and output modalities to choose from. Sometimes, the messages a system needs to communicate are often more appropriately conveyed in abbreviated or metaphorical terms. Context can often best dictate modality and metaphor.

As music is a language of expressivity, HCI has direct bearing on the successful design of musical interfaces. Traditional acoustic instruments act as both interface and transducer for expressivity. in computer-generated music and sound, however, the interface can be more ambiguous. Some researchers feel HCI should provide for a discrete mapping between gestural control by a musician and the expressive output of that instrument. Fels describes the relationship in terms of intimacy [4]. When the gestural mapping is transparent, a high degree of intimacy results, allowing a musician to embody his or her instrument, whether it be acoustic or electronic. Sometimes, though, increased expressivity can be generated out of the ambiguity and inefficiencies inherent in an interface [5]. Good design should, however, incorporate ambiguity in a meaningful way. Context may decide which components, ambiguous or unambiguous, are best for a particular application.

3 METAPHOR

Metaphor is a means of describing one idea through another idea. Perhaps the most familiar use of metaphor in computing is the abstract appropriation of the term "desktop" to the graphic work environment of a computer. The word, generally regarded as a breakthrough in graphical user interface design [6], carries so much meaning that it has come to embody the idea that it represents. In terms of interface design, good use of metaphor can perform a channel coding function of sorts, especially when considering the great amount of information that must pass between human and computer.

Metaphors can be transmitted through any modality. Like graphic metaphor, sonic metaphors can represent in a short temporal space an idea that occurs over an extended period of time, or vice versa. They, similar to lookup tables utilized for data reduction, can be used as symbols for frequently occurring events or actions. Expressivity and creativity can easily be hindered by excess verbiage; good metaphor in interface design limits the amount of redundant information passing between computer and human. Additionally, metaphor can be used to transfer data from one domain (ie visual) to another (auditory). For the purposes of this paper, auditory metaphor will be examined in terms of auditory icons and data sonification.

3.1 Icon

Auditory icons minimize informational overload in a visual display by conveying messages that don't necessarily need to be seen. An auditory icon (sometimes called an earcon) describes actions that are occurring in the background, warns users about events, or alerts a person to shift their attention from one display to another. Spatial auditory icons are powerful tools and have shown to be a good substitute for visual feedback for the visually impaired [7]. An auditory icon can be abstract or literal, depending on its application. In [8], timbral modifications of a sound were selected as a good indication of tactile roughness (abstract). In another experiment, auditory icons were used to signify invisible events occurring in a simulated beverage manufacturing plant (literal). Because the auditory icons placed the users in the same auditory space, they were useful for enabling two users to cooperate while occupying independent virtual spaces [9].

3.2 Sonification

Sonification is the representation of a data stream through sound. Sonification is advantageous in situations where we may use our ears to detect subtle changes that are otherwise difficult to perceive. Our ears' temporal acuity and sensitivity to changes in level and pitch make sonification an ideal method for displaying complex data streams. The real-time characteristics of sonification and the fact that spatial audio can be localized to almost anywhere in a space means that information that would otherwise saturate a visual display can be successfully transmitted to a human through less restrictive means. Stock market, weather, and EEG brain wave data have been sonified for artistic and non-artistic means in non-interactive ways. That is, the displays are one-way. But interactivity and manipulation lend themselves to sonification, as well [10]. For instance, the actions of a human being can be sonified in such a way as to transmit meaningful information about those actions. In "The Music Without" a violinist's bowing actions are sonified while playing, resulting in two sonifications: one of violin music, and another of violin operation [10]. A challenge of sonification is that it must successfully coexist with other, uncorrelated, sounds in an environment.

4 MODALITY

Humans can interface with computers through a variety of modalities. Traditionally, visual display has been a favored output device of computing systems, while keyboards and mice act as common input devices. These are rather insufficient at communicating complex human information to a computer. Neither of these modes of communication is similar to communication occurring naturally in the human-to-human world.[1] These traditional input/output methods encourage a stiff, uncreative relationship between computer and user. Additionally, data transmission between humans and computers can be rather clumsy, as we have adapted our language and behaviors to that of machines.[2] Indeed, not all human signals are vital to communicating important information, but the great majority of human communication channels are untapped when it comes to computers. These include gestures, vocalizations, haptic signals, and others. Pantic argues that computers must learn to interpret a human's "affective state", assimilated from across multiple modalities. Comprehension of affective state is described as "the core of social and emotional intelligence," having the capacity to correlate a discrete emotion or social signal [3]. Currently, human affective state is poorly taken into account in most interface design.

[1] Except when visual displays are used to display human images on computers.

[2] Of course, with enough practice, we can learn the language of keyboards, mice, and graphic representations. However, these devices don't adequately utilize the communicative channels that are the most developed for humans.

4.1 Visual

While the visual mode of communication has been exploited in computers, there is still much content not being capitalized. On the computer output side, computer images should sometimes exhibit their displays in their own, virtual, ”affective states”—in terms that humans instantly recognize. Graphic representations of human gesture could go far in communicating important content, simulating the way humans communicate with one another.

Computer input should likewise have the capacity to read our gestural communications. Eye and face tracking will enable us to direct computing environments in more humanistic ways and overcome mode-specific noise. For instance, Kauckėnas *et al.* describe a video system that detects face and mouth movements. The additional input modality augments a speech recognition system, as visual input isn't subject to acoustic noise and can help clarify sonically ambiguous consonants [11].

In some instances, visual display must accommodate environments that are saturated in the audio domain. Zinman and Donath [12] describe a visual display that is intended to handle issues of scalability for a persistent audio environment on mobile telephones. The phones use a platform that supports multiple user chats through audio voice messaging. The visual display is designed to iconically represent a large community of users and provide easy navigability through mobile phone chats.

Visual display is taken for granted, yet isn't always used economically. There is limited real estate available when transferring data in the visual domain. Furthermore, visual displays can place high cognitive load and demand undue attentional energy from a user. It is important to be selective in choosing whether visual display is the best modality for a particular function.

4.2 Haptic

Haptic input is useful to humans, as it allows us to apply our knowledge of the physical world to a virtual one. Combined with haptic feedback, haptic input can be a simple means for users to assign values to a virtual environment with greater accuracy than is possible with visual or auditory input. Mild haptic feedback can create constraints in a system that add an additional dimension to a virtual environment. In one study, weak attractive forces were found to provide substantive information about a virtual spatial enviroment [13].

Indeed, because our understanding of the world is primarily physical, [14] suggest that haptic interface design should be modeled around a system of constraints and tokens. That is, interfaces should comprise physical objects representing digital information within a system of physical constraints. This methodological approach toward haptic input and feedback, they argue, shifts cognitive load away from objects being manipulated and provides an increased sense of kinesthetic feedback.

The primary input modality of traditional musical instruments is, in nearly all cases, haptic. Nontraditional instruments can benefit greatly from including haptic modality, as haptic input is a simple way to capture a large range of human gestures. Additionally, the feedback from a haptic output can improve the overall expressivity of an instrument.

4.3 Auditory

Auditory display is useful in scenarios where visual information overload is a problem or in environments where auditory signals are better understood than visual. Providing sound as a primary display during teaching exercises for school children has been shown as a successful teaching aid. The immersive and social characteristics of auditory display integrated children into exercises at hand [15].

The music making process can be enhanced when humans and computers use common auditory signals to communicate events or processes to each other. Similar to common human-to-human communication during musical performances, musicians can use expressive, contextually appropriate, signals that notify an interface to handle his or her input differently. Likewise, similar auditory feedback can be output by a system during a performance to signal background processes. A main advantage of auditory modality over others is that it does not bind a user to a specific location within a space; a user does not need to be within sight lines of a computer screen or within reach of a haptic input device.

As mentioned above, auditory display faces the challenge that it will almost always have to coexist with ambient noise and preexisting sounds. Self-organizing auditory display, however, would intelligently place feedback within vacant sonic spaces so that it can maximize its usefulness within an environment [16].

4.4 Gesture

Gesture may not be considered a discrete input/output modality, as it relies on all modalities to communicate a wide breadth of information. Gestural control is, however, extremely important to modern interface design. Gesture is one of the basic building blocks of human communication. Furthermore, it isn't restricted by the syntax of spoken language. Gesture is incorporated into nearly every human movement or action. Mining useful gestures from humans is complicated and requires a high level of intelligence. Ideal computer systems will have the intelligence to observe human gesture and, also, output data in gesturally significant ways.

Countless methods of gestural control have been implemented or suggested. Examples of direct approaches include sound sculpting [17] and gestural control of audio effects [18] with the hands. Indirect gestural control can involve sonification of gesture such as, iterated earlier in this paper, the bowing movements of a violinist [10] for further manipulation and processing. The success of gestural control is partly dependent on intuitive mapping between control parameters and output parameters. Good implementation of gestural input and display will, arguably, close the gap between humans and computers.

4.5 Multimodality

Multimodal inputs and displays combine several channels of information in ways that exploit the strengths of each. For instance, the audiovisual integration implemented in the speech recognition software mentioned earlier in this paper [11] combines the visual and auditory modalities to overcome deficiencies in each. Similarly, the virtual beverage manufacturing experiment [9] proved that selective use of auditory and visual modalities can improve task completion in a cooperative challenge. Or, the combination of haptic and auditory modalities can give more detailed information about a texture than either modality on it's own [8]. Magnusson *et al.* have shown that mild haptic constraints, coupled with spatial auditory icons, can construct a robust, navigable virtual space [13].

5 CONTEXTUAL AWARENESS

Intelligent awareness of a user's context greatly enhances the functionality of a computing system. Changing contexts can have important effects on a user's performance of certain tasks, and multiple contexts can interact with each other in unpredictable ways [19]. Context can be drawn from an infinite number of sources; discerning which contextual information is pertinent to a task at hand is challenging. But if context is harvested and filtered properly, it is an undeniably strong force for good interfacing.

Some researchers believe that context should first be defined by a specific user; a system can then adapt from there [20]. They have emphasized that balance must be maintained between human and software control in contextually aware applications. End-user programming, where users ultimately configure a system to be aware and in control of elements that fit their needs, is advocated.

Contextual awareness is especially popular in the mobile computing fields, as the nature of mobile computing involves ever-changing contexts. Ma *et al.* have reported exciting developments in automatic acoustic environment classification [21]. Designed for mobile devices, their software can classify an acoustic environment with surprisingly high accuracy before a human can. Earlier work [22] reports a device that transmits information about the weather in a context-sensitive manner using acoustic environment classification. For music, context can be harvested in a number of ways. Information about performance space, audience type, time of day, or temporal location within a composition can all be used to contribute to a musical experience.

Context-aware systems must be sensitive to the orientation of focus of its users. Intelligent contextually aware systems should delicately manage a person's focus between informational domains and modalities. Context-aware audio systems have great applicability in audio systems, especially for healthy data sonification or auditory display. An awareness of acoustic environment can help a system to choose when to use an auditory display, which metaphors to apply or how that display is utilized, keeping the acoustical ecology of an environment in balance.

5.1 Focal management

When presented the multiple modalities for information input and output, it is useful for computing systems to manage their users' attention. Information-rich applications can help to place more important information at the forefront of a user's awareness. For instance, in one study, the differences between the effect of a high-quality video feed and a low-grade graphical representation upon task completion were studied [23]. It was found that video feeds were better at encouraging conversational turn-taking from the participants. However, graphical representations of participants demanded less focal energy and allowed participants to better complete focus-oriented tasks during the conference. Contextually aware systems can determine where a user's attention should be directed and manage modal output based upon those decisions.

Sanderson *et al.* tackle the issue of moving auditory displays into and out of focal awareness in [24]. They suggest that if some phases of Cognitive Work Analysis (CWA) is applied and attentional mapping is added, successful management of focal awareness can be achieved. They comment that once focal awareness is shifted to a new modality, other modalities are often largely unattended to. They emphasize the necessity of integrating auditory display from the beginning of interface design, rather than as an add-on, so that focal management is integrated into its basic design.

5.2 Ecology

Acoustic ecology in interface design infers humans interacting with each other and with sounds in an acoustic environment. As mentioned earlier in this paper, the intelligibility of an auditory display will almost always be challenged by external sounds. Sound can be more meaningful as an information source if it is appropriately placed within an ecology of sounds. Self-organization is one solution [16].

Auditory display can be placed within the acoustic ecology of any information-rich environment, such as a workplace. Scientists in Sweden conducted an experiment on the informational quality of environmental sounds from a factory [25]. They found that workers in a factory derive a great deal of auditory information from their work environment and speculated that similar acoustic ecologies with high informational worthiness can be created in a virtual space.

Similarly, musical performance often incorporates a highly sophisticated acoustic ecology between sonic layers, performers, and audience. New, ecologically aware instrument design can lead to automated placement of sound performance within congruous pockets of a sonic texture.

6 MUSICAL AND AUDITORY APPLICATIONS

New interfaces for musical creativity and auditory display are increasingly realizing the potential of HCI concepts outlined in this paper. Just a few examples, with some of their strengths and shortcomings, are outlined below.

6.1 *The ReacTable*

The ReacTable [26], a tabletop musical interface for multiple or solo performers, exhibits a wonderful integration of multimodality for musical performance. To play the instrument, objects are moved along a flat, horizontal surface. The objects are in the forms of pucks with simple, geometric shapes. Each puck shape signifies a unique instrument object and performs a discrete process. The system software identifies the pucks and tracks their movement across the table. A visual display illuminates the table from underneath, providing the performers with important informational feedback about the status of the system.

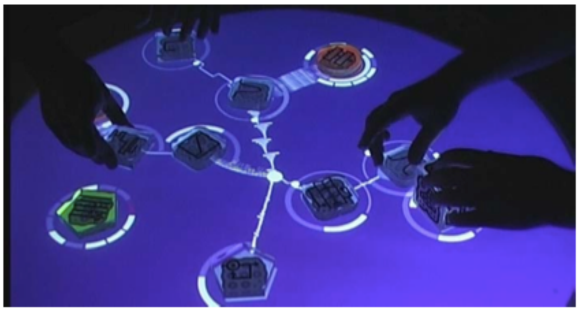

Figure 1: The *ReacTable*
[26]

Musicians perform through haptic input and respond to visual, auditory, and passive haptic feedback (Figure 1). Haptic manipulation of the geometric objects allows for precise positioning upon the table. After a short learning procedure, users learn to associate objects with their shapes, introducing metaphor into the system. The musical interface allows for a high degree of interaction between users and instrument. Additionally, the physical constraints of the surface and discrete objects conform to Ullmer *et al.*'s theory of constraints [14]. The system, however multimodal, doesn't provide for much gestural control and could perhaps benefit from such input. In either case, however, the instrument promotes interaction between users and computer and encourages creativity and expressivity.

6.2 *Lady's Glove*

The *Lady's Glove* [5], created by Laetitia Sonami (pictured in with the glove in Figure 2), is a musical controller consisting of a glove worn on the left hand and additional sensors placed at various places on the body. Incorporating bend, magnetic, light and sonar sensors, along with accelerometers, the interface allows excellent gestural control. Input modalities are haptic and gestural, and display is auditory. The entire interface can be considered metaphorical, as hand gestures are translated through a glove that is worn on the hand. The glove's mappings to specific sounds are ambiguous, but the sophistication of the design leads to some level of control over those ambiguities. Creativity is fostered, as the input modalities do not bind the user in one place.

Figure 2: The *Lady's Glove*
Source:
http://emfinstitute.emf.org/exhibits/ladysglove.html

6.3 The *MusiCube*

The *MusiCube* [27] is a music player designed to package haptic, visual, and auditory modalities in a simple music output interface (Figure 3). Rather than using a display screen for informational status, flashing colors provide a metaphor for the functions that the device is performing, such as system status and volume level. There are other metaphorical implementations in the device, as well. Shaking it causes it to enter shuffle mode. Playlists are color-coded; placing a particular color face up activates its associated playlist. The texture of the device allows a user to interact with it in a more tactile way than an ordinary player, augmenting a sense of control for the user. A synthesized speech generator announces the playlist, song title, function, and volume level of the device, augmenting visual feedback with auditory feedback. The design of the player attempts, at every instant, to facilitate interaction through feedback and metaphor. In [27], there is no indication, however, as to whether the auditory feedback of the player fits ecologically with the device's music output functions.

Figure 3: The *MusiCube*
[27]

7 A MODEL FOR INTERFACE DESIGN

Modern interface design for musical and auditory applications should exhibit current developments in HCI. The metaphors that are used should be easily recognizable, intuitive, or easily learnable. Channel-coding of redundant information should be exercised through good metaphor and sonification. In all parts of the design, a simplification of the communicative process between human and computer should be encouraged.

Multimodality of input and output should be employed skillfully—to inhibit informational overload, unbind a user from his or her workstation or device, and to improve communication channels between humans and computers. Modalities should make ecological sense and should be easily adaptable to changing circumstance.

Gestural awareness and intelligence should be a key component of interface design. Gestural information should be filtered properly and mapped well to an interface. Mappings may not always be unambiguous, yet they should leave a user with a strong impression of control.

Interfaces should be aware of context and be able to apply changing modality and metaphor to changing context. This needs to happen in such a way so as users don't become disoriented or feel a loss of control over the interface. Context should be user-definable and re-definable. Also, focal awareness must be successfully managed, handling modal salience in an intelligent manner.

Interfaces should break down communicative barriers between humans and machines. They should understand our behavior, react to changing context, and display information in data-reduced, yet meaningful terms. Above all, interfaces should foster creativity and become an extension of the human process.

8 CONCLUSION

This paper has shown how metaphor, modality, and context can be applied towards Human-Computer Interaction and better interface design for musical and auditory applications. Brief reviews and summaries of HCI, metaphor, modality, and contextual awareness were given. A few applications incorporating some of these concepts were presented, and a model for improved interface design was suggested.

References

[1] M. G. Petersen, K. H. Madsen, and A. Kjær, "The usability of everyday technology: emerging and fading opportunities," *ACM Trans. Comput.-Hum. Interact.*, vol. 9, no. 2, pp. 74–105, 2002.

[2] M. Pantic, A. Pentland, A. Nijholt, and T. Huang, "Human computing and machine understanding of human behavior: a survey," in *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, (New York, NY, USA), pp. 239–248, ACM Press, 2006.

[3] M. Pantic, N. Sebe, J. F. Cohn, and T. Huang, "Affective multimodal human-computer interaction," in *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, (New York, NY, USA), pp. 669–676, ACM Press, 2005.

[4] S. Fels, "Designing for intimacy: Creating new interfaces for musical expression," *Proceedings of the IEEE*, vol. 92, no. 4, pp. 672–685, 2005.

[5] L. Sonami, "The lady's glove." Lecture at New York University, April 18 2007.

[6] A. F. Blackwell, "The reification of metaphor as a design tool," *ACM Trans. Comput.-Hum. Interact.*, vol. 13, no. 4, pp. 490–530, 2006.

[7] A. B. Barreto, J. A. Jacko, and P. Hugh, "Impact of spatial auditory feedback on the efficiency of iconic human-computer interfaces under conditions of visual impairment," *Comput. Hum. Behav.*, vol. 23, no. 3, pp. 1211–1231, 2007.

[8] E. E. Hoggan and S. A. Brewster, "Crossmodal icons for information display," in *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, (New York, NY, USA), pp. 857–862, ACM Press, 2006.

[9] W. W. Gaver, R. B. Smith, and T. O'Shea, "Effective sounds in complex systems: the arkola simulation," in *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, (New York, NY, USA), pp. 85–90, ACM Press, 1991.

[10] K. Beilharz, "Gesture-controlled interaction with aesthetic information sonification," in *IE2005: Proceedings of the second Australasian conference on Interactive entertainment*, (Sydney, Australia, Australia), pp. 11–18, Creativity & Cognition Studios Press, 2005.

[11] J. Kaukėnasnas, G. Navickas, and L. Telksnys, "Human-computer audiovisual interface," *Information Technology and Control*, vol. 35, no. 2, pp. 87–93, 2006.

[12] A. Zinman and J. Donath, "Navigating persistent audio," in *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, (New York, NY, USA), pp. 1607–1612, ACM Press, 2006.

[13] C. Magnusson, H. Danielsson, and K. Rassmus-Gröhn, "Non visual haptic audio tools for virtual environments.," in *HAID* (D. K. McGookin and S. A. Brewster, eds.), vol. 4129 of *Lecture Notes in Computer Science*, pp. 111–120, Springer, 2006.

[14] B. Ullmer, H. Ishii, and R. J. K. Jacob, "Token+constraint systems for tangible interaction with digital information," *ACM Trans. Comput.-Hum. Interact.*, vol. 12, no. 1, pp. 81–118, 2005.

[15] M. Droumeva, A. Antle, and R. Wakkary, "Exploring ambient sound techniques in the design of responsive environments for children," in *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, (New York, NY, USA), pp. 171–178, ACM Press, 2007.

[16] D. Brock, J. A. Ballas, and B. McClimens, "Perceptual issues for the use of 3d auditory displays in operational environments," in *ISICT '03: Proceedings of the 1st international symposium on Information and communication technologies*, pp. 445–448, Trinity College Dublin, 2003.

[17] S. Hashimoto and H. Sawada, "A grasping device to sense hand gesture for expressive sound generation," *Journal of New Music Research*, vol. 34, pp. 115–123, Mar 2005.

[18] T. Hermann, S. Paschalidou, D. Beckmann, and H. Ritter, "Gestural interactions for multi-parameter audio control and audification," in *Gesture Workshop* (S. Gibet, N. Courty, and J.-F. Kamp, eds.), vol. 3881 of *Lecture Notes in Computer Science*, pp. 335–338, Springer, 2005.

[19] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears, "Capturing the effects of context on human performance in mobile computing systems," *Personal Ubiquitous Comput.*, vol. 11, no. 2, pp. 81–96, 2007.

[20] B. Hardian, "Middleware support for transparency and user control in context-aware systems," in *MDS '06: Proceedings of the 3rd international Middleware doctoral symposium*, (New York, NY, USA), p. 4, ACM Press, 2006.

[21] L. Ma, B. Milner, and D. Smith, "Acoustic environment classification," *ACM Trans. Speech Lang. Process.*, vol. 3, no. 2, pp. 1–22, 2006.

[22] D. Smith, L. Ma, and N. Ryan, "Acoustic environment as an indicator of social and physical context," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 241–254, 2006.

[23] E. L. Clayes and A. H. Anderson, "Real faces and robot faces: The effects of representation on computer-mediated communication," *International Journal of Human-Computer Studies*, vol. In Press, Corrected Proof, pp. –, 2007.

[24] P. Sanderson, J. Anderson, and M. Watson, "Extending ecological interface design to auditory displays," in *OZCHI '00: Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia*, pp. 259–266, CSIRO, 2000.

[25] P. Alexanderson and K. Tollmar, "Being and mixing: designing interactive soundscapes," in *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*, (New York, NY, USA), pp. 252–261, ACM Press, 2006.

[26] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner, "The reactable: exploring the synergy between live music performance and tabletop tangible interfaces," in *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, (New York, NY, USA), pp. 139–146, ACM Press, 2007.

[27] M. B. Alonso and V. Keyson, "Musiccube: a physical experience with digital music," *Personal Ubiquitous Comput.*, vol. 10, no. 2, pp. 163–165, 2006.