Gesture-Based Interactions with Virtually Embodied Wearable Computer Software Processes Competing for User Attention

Nicolas Plouznikoff, Alexandre Plouznikoff, Michel Desmarais, and Jean-Marc Robert

***Abstract*—One current limitation of wearable computers preventing their use as our everyday companions is directly linked to the interaction mechanisms employed for information selection: interactions with wearable computers are simply not integrated enough into the flow of real world user actions. This paper presents a new interaction paradigm where a user can interact with competing software processes embodied in his/her environment through tangible and meaningful real-world actions. After reminding the need for seamless human-wearable computer interactions, we describe this novel concept, the benefits of using natural actions in a mobile setting and the hardware and software architecture of our current prototype. A preliminary user study shows that this paradigm is well received and even preferred over more conventional, and slower, interaction mechanisms.**

I. Introduction

Wearable computing still presents numerous challenges [1]: hardware miniaturization, garment integration, mobile augmented reality, context awareness, mass market and socially acceptable applications, etc. However, one of the main challenges remains human - wearable computer interfaces.

In general, a wearable computer's role is to directly or indirectly assist a mobile user carrying out a real-world task in an environment where human - wearable computer interactions are not the primary focus [2]. For these interactions to seamlessly blend into the task at hand, and avoid costly transitions between the real world (the task realm) and the virtual world (the wearable computer realm), suitable interaction paradigms and low attention input / output mechanisms are needed. This is also true when the wearable computer is only indirectly supporting a task (i.e. when offloading the user by taking care of various background tasks). To allow a user to intuitively and transparently interact with wearable computer software processes and/or agents competing for his/her attention, we investigate the use of natural human behaviors and actions as input methods.

This work was supported in part by two NSERC grants (National Science and Engineering Research Council of Canada) and one FQRNT grant (Quebec's Research Fund on Nature and Technologies).

All authors are with the Human-Machine Interactions Laboratory (LIHM) at École Polytechnique de Montréal, Montreal, QC, Canada, and can be reached by email: {nicolas.plouznikoff, alexandre.plouznikoff, michel.desmarais, jean-marc.robert}@polymtl.ca.

II. Related Work

A large body of literature exists on wearable computers interfaces. It can roughly be divided into research drawing upon explicit interactions (i.e. utilizing dedicated peripherals or actions), [3, 4] for example, and research focusing on implicit interactions (i.e. trying to exploit context), [5] for example.

Research related to natural interfaces [6] is mostly devoted to supporting specific real-world tasks, either through voice commands, gestures, object manipulations, etc. or a combination of the previous. For example, ReachMedia [7], which uses RFID tags and discrete wrist gestures for input and audio for output, enables socially acceptable hands-free and eyes-free explicit interactions with augmented physical objects (books). Such an interaction paradigm, where real-world entities become the interface to the wearable computer, allows a user to transparently access his/her wearable computer's functionalities through the real world [8].

More mature research already investigated how to lessen a user's perceptual and cognitive loads, either by minimizing the quantity of information presented, using just-in-time information presentation [9], or by integrating this information in the environment, using augmented reality techniques [10].

III. Interacting with Virtually Embodied Software Processes

Instead of focusing on assisting single monolithic tasks, this paper explores an interaction paradigm suitable for multi-task support. As opposed to tangible user interfaces [11], the goal isn't to physically embody virtual data but rather to virtually embody in the environment software processes running on the wearable computer in order to let the user interact with them as naturally as possible.

Imagine for example that, while you are walking about or driving your car, three events occur: a new email arrives in your inbox (communication related information), a snow storm is announced (weather information) and the time comes when you need to take a specific medication (medical information). All of those three pieces of information could potentially be of interest to you. How then can you express your interest (or lack thereof) for each type of information and eventually access the related information through your wearable computer in an unobtrusive way? What if you could naturally brush away (or maybe kick away) the information you do not need and grasp the one you're interested in?

After creating dedicated software processes (or more "intelligent" software agents), with which each type of information is associated, and embodying these software processes into the user's environment using virtual avatars in charge of attracting user attention, we turn to natural actions (grasping, touching, brushing, stomping, kicking, etc.) to interact with those processes. Here, augmented reality provides a way to integrate these processes into the environment and to avoid back and forth transitions between the real and the virtual world. Instead of forcing the user to interact with these processes/agents through specific devices (conventional computer peripherals for example) or artificial spaces (graphical user interfaces for example), the wearable computer functionalities become accessible through real-world actions and the wearable computer interface is perceived as emanating from the user's immediate surroundings.

The key advantage of this interaction paradigm is that, for a user, these kinds of explicit physical actions are intuitively representative of meaningful information manipulation. By exploiting this naturally occurring phenomenon, a user can spontaneously recall and engage in actions to manage the presented information.

IV. Hardware Implementation

Fig. 1 depicts the wearable computing platform used to test our interaction paradigm. The core system, integrated inside a user-worn vest, is built around a PC-104+ main module equipped with a Transmeta Crusoe 1.0GHz processor, 256MB of RAM, a 80GB mobile HDD and an add-on frame grabber. A MicroOptical SV6 monocular opaque display (640x480 pixels at 60Hz, 18 bits colors) was used as the display device to present the user with an altered view of the world.

Various sensors are used to detect and recognize the user's

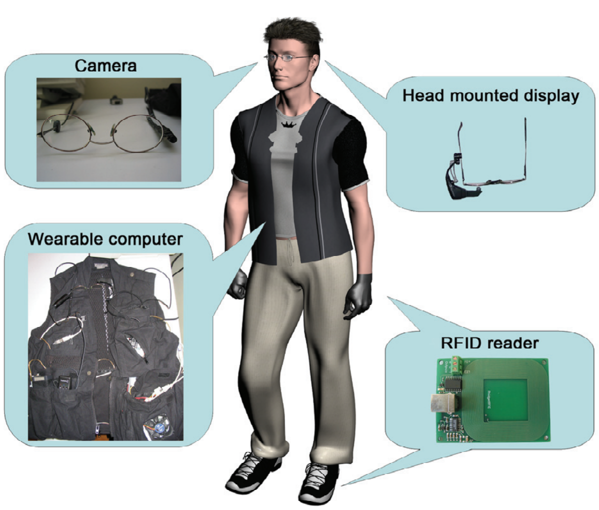

Fig. 1. Wearable computing platform.

actions. A low power miniature video camera with a color CCD was mounted on the side of the user's glasses (or transparent safety glasses when the user wasn't already wearing glasses) in order to acquire live video of the user's point-of-view. A Phidgets RFID reader (125KHz tag frequency, 30Hz read update rate, 2 to 3 inches maximum read distance) was integrated inside a glove worn by the user and was used to collect proximity/contact information. We also fitted a RFID reader into the sole of a modified shoe to investigate further interaction possibilities (see section X).

V. Software Implementation

In order to implement our interactions techniques with software processes we need 1) to virtually embody those processes; 2) to carefully assign a virtual meaning to natural user actions and 3) to detect and recognize those natural actions.

A. *Software Process Embodiment*

For our experiments, we considered three types of information: communication related information, weather related information and medical information. Each type of information is handled by a corresponding software process. A unique and representative 3D avatar was created to represent each software process. Avatars used in the experiments described below where stationary. However, they can easily be altered and/or animated according to the subjective importance of the information they are trying to communicate or even to directly reveal part of this information. Simple color-based alterations were used in the first experiment (see section VII).

To virtually embody a software process in the user's surroundings, our C++ application relies on the ARToolkit library and associated visual markers placed in our test environment. Thanks to those placeholders, avatars can be registered in the environment and seen in the altered view presented to the user. When an alert has to be presented to the user, the first vacant placeholder found in the user's field of view is used. The user's hand is also tagged with such a visual marker (to register a picked-up avatar on the user's hand, see Fig. 2).

B. *Interaction Design*

Our goal is to let a user interact naturally with each avatar to access or manipulate the corresponding software process. To detect proximity/contact with an avatar, a unique RFID tag was inserted below each visual marker. In order for interactions to be intuitive, a virtual meaning was then assigned to two very specific tangible user actions for the following experiments: picking-up and brushing away.

Picking-up signifies interest for the software process embodied by the avatar. By touching a visual marker, the avatar is transferred to the user's hand and the associated software process is instructed to deliver information as long as the avatar is in the user's field of view.

Fig. 2. A simple gesture-based interaction: picking-up the weather avatar on the floor to indicate interest for this software process.

Brushing away an avatar indicates a lack of interest for the associated software process. It instructs the software process to postpone the alert for a preset period of time (refractory period). What's interesting in this case is that we could eventually distinguish between various "brush types": a quick brush would mean that the alert is of no use to the user and would discard the information, whereas a slower brush would simply delay the alert according to the intensity of the brush.

C. Heuristic Interaction Recognition

Based on initial observations, heuristic rules were empirically devised to recognize user actions.

Picking-up is initiated by a hand proximity/contact (using the instrumented glove) with a marker for more than 500ms. The pick-up lasts until the hand marker drops from the user's view for more than 2000 consecutive ms.

Brushing away is characterized by a hand proximity/contact with a marker for less than 500ms, the intensity of the brush being inversely proportional to the proximity/contact time.

VI. Compared Interaction Methods

To compare and evaluate the benefits of our gesture-based approach versus more conventional interaction methods, we devised two preliminary experiments. The first one tried to quantify which of these methods led to the smoothest transitions between the real and the virtual when they both compete for the user's attention; while the second experiment tried to assess the efficiency of gesture-based interactions.

We initially considered four conventional interaction methods for our comparison (see Table I), where the user had to achieve the same tasks via gesture-based interactions and with those conventional methods. For those methods we used a Twiddler2 (one handed chording-keyboard with integrated trackpoint pointing device geared towards mobile users) as the input peripheral and displayed avatars as giant 2D icons in front of a white background on the head mounted display. The Twiddler2 was basically used as a mobile pointing device and no real typing was required. For example, using method C2, users were able to point an icon/avatar and interact with it by pressing one of two buttons (key "A" to postpone the associated alert and key "E" to view it; those keys being the easiest to reach).

TABLE I
INVESTIGATED INTERACTION METHODS

Method (abbrev.)	**Positioning**	**Action selection**
Gesture-based (G)	Hand positioning	Meaningful hand gestures
Conventional 1 (C1)	None	Shortcut keys using the built-in chording mode (3 avatars x 2 actions = 6 combinations)
Conventional 2 (C2)	Trackpoint	Shortcut keys (2 actions = 2 "buttons" used)
Conventional 3 (C3)	Trackpoint	Contextual menu (1 key used)
Conventional 4 (C4)	Trackpoint	Pointer gestures

VII. Experiment 1: Single Asynchronous Alert

To assess the reaction time and the recall easiness of our gesture-based interaction mechanisms for relatively unprepared users, we performed a simple experiment. We enrolled 8 users aged 25 to 39, with no known disabilities, for a two-part experiment followed by a debriefing. Of those 8 users only 7 showed up for the tests. Before the start of the experiment, we introduced them to the wearable computer as well as to our 3 avatars and their respective roles. We also explained how each interaction method worked and each user had 5 minutes before each part of the experiment to practice the interaction method tested.

A. Experiment

We individually sat each user at a specially prepared desk on which was placed a single visual marker / placeholder. We then gave them a Rubik's cube to play with (to occupy both their hands and mind) and told them to try to solve it. We also told them that during the next 5 minutes an alert would pop-up, either as an avatar registered with the placeholder (for the first run of the experiment) or as a 2D image (for the second run of the experiment). As soon as the alert popped-up, the user had to either "pick it up / accept it" or "brush it away / dismiss it". The action to be performed was chosen randomly each time and was indicated by the tint of the modified avatar: a green tint meaning "pick it up / accept it" and a red tint meaning "brush it away / dismiss it". The first part of the experiment relied on gesture-based interaction whereas the second part relied on interaction method C2, the conventional point and click (however the user only had to click the correct button without moving the pointer since there was only one avatar presented full screen). The time taken to perform the correct action (i.e. to interact correctly with the avatar) was recorded, as well as the number of mistakes made (i.e. the number of wrong interactions even though errors had no effect on the avatar).

B. Results

The times recorded (in seconds, see Fig. 3) are an indicator of the time taken to transition between the real and the virtual, and to recall the correct interaction mechanism to perform the requested action.

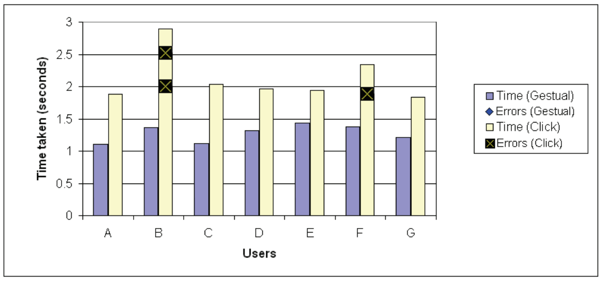

Fig. 3. Time taken and errors made to complete a "dismiss or accept" task using gesture-based interactions and interaction method C2.

The time to complete the task is on average lower when using gesture-based interactions (M=1.28, SD=0.13) compared to method C2 (M=2.13, SD=0.37). We defined the null hypothesis (H_0) as $\mu_G \geq \mu_{C2}$ and the alternative hypothesis (H_a) as $\mu_G < \mu_{C2}$; μ_x being the mean time taken to complete the task using interaction method x. Assuming the paired time differences to be normally distributed (even though the sample size is small, a Shapiro-Wilk test gave no sufficient evidence against normality: W=0.897, p=0.315, α=0.05), a one-tailed paired samples t-test led us to reject hypothesis H_0 and revealed that the lower mean times recorded for interaction method G compared to method C2 were statistically significant: t(7)=-6.854, p<0.001, α=0.05; the mean difference being -0.85 and the 95% confidence interval for the difference -1.16 to -0.55. In addition, we observed during the experiment that the time taken to reach and grab the Twiddler was often the main contributor to the total time taken to perform the required action with method C2. This means that even if we had tested another conventional interaction method (C1, C3, C4), the recorded times would probably have been just as high.

A few errors occurred with method C2 whereas no errors were recorded when using gesture-based interactions (see Fig. 3). However, a two-tailed Wilcoxon paired samples signed rank T test failed to reveal a statistically significant difference between the underlying distributions of the number of recorded errors for interaction method G (M=0.00, SD=0.00) and method C2 (M=0.43, SD=0.79): T=0, p=0.180, α=0.05.

VIII. Experiment 2: Multiple Concurrent Alerts

To assess the efficiency of gesture-based interactions versus more conventional methods when several processes are in competition for user attention we performed a second three-part experiment. The same 7 users were used for this experiment and, as in the first one, we allowed them to practice with each interaction method for 5 minutes beforehand.

A. Experiment

The first part of the experiment focused on gesture-based interactions. Each subject was placed in front of 3 placeholders (one for each avatar) evenly spaced on a desk. There, each subject was asked to react to 3 simultaneous alerts: first, the user had to postpone the weather alert, then postpone the email alert and finally, obtain more details about the medical alert. Here again, the time to complete the task (i.e. to interact correctly with the three avatars) was recorded. Wrong interactions were also recorded but had no effect on the avatars (for example the medical alert could not be accidentally discarded by brushing it away).

We repeated the same procedure for two conventional interaction methods with which the users had to achieve the same task. We chose to test methods C1 and C2 for two reasons: first the positioning is different, and secondly because contextual menus and pointer gestures only had the potential to increase the number of manipulations required and thus increase the time taken to complete the task (even though they could potentially decrease the number of errors). For method C1, users had to recall and use the key combination which was associated with the correct avatar and the correct action (6 possible combinations). For method C2, users simply had to point the avatars and recall and use the correct key to be pressed (2 possible keys).

B. Results

Compared to the first experiment, users are here ready to interact with the avatars. The recorded times (in seconds, see Fig. 4) are an indicator of the efficiency of the interaction mechanism for this task.

The time to complete the task is on average lower when using gesture-based interactions (M=1.95, SD=0.07) compared to interaction method C1 (M=5.44, SD=2.25) or C2 (M=4.66, SD=0.59). We assumed the recorded times for each method to be normally distributed since Shapiro-Wilk tests gave no sufficient evidence against normality for interaction method G (W=0.980, p=0.958, α=0.05), method C1 (W=0.888, p=0.263, α=0.05) and method C2 (W=0.842, p=0.104, α=0.05). A one-way repeated measures ANOVA using Greenhouse-Geisser degrees of freedom correction (Mauchly's test indicated that the assumption of sphericity had been violated: χ^2=8.44, p=0.015, α=0.05, ε=0.551) led to the rejection of H_0 ("there are no significant difference between the mean times for interaction method G, C1 and C2") and revealed a statistically significant difference between the mean times being compared: F(1.102,6.612)=14.037, p=0.007, α=0.05. Post-hoc tests

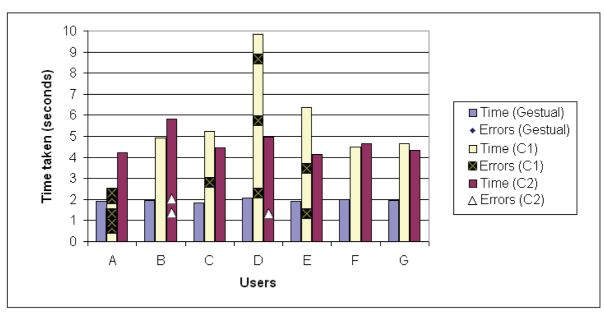

Fig. 4. Time taken and errors made when trying to complete the task using gesture-based interactions, interaction method C1 and interaction method C2.

(repeated-measures t-tests with Bonferroni correction) revealed that the participants were significantly faster when using interaction method G compared to method C1 ($p=0.018$, $\alpha=0.05$; the mean difference being -3.49 and the 95% confidence interval for the difference -6.24 to -0.74) as well as method G compared to method C2 ($p<0.001$, $\alpha=0.05$; the mean difference being -2.71 and the 95% confidence interval for the difference -3.42 to -2.00). No significant statistical difference existed between the mean times for method C1 and C2 ($p=1.000$, $\alpha=0.05$). The statistically significant increased performance for gesture-based interactions could be attributed, for novice users, to the intuitiveness of gesture-based interactions with registered avatars compared to other interaction methods. User feedback collected during our debriefing survey corroborated this hypothesis: all participants (7 out of 7) found method C2 easier to use than method C1 but participants still preferred (6 out of 7) gesture-based interactions over both. This preference could however be tied to the novelty factor.

For method C1, users 1) couldn't remember which button was associated to which action (high number of errors) and 2) had a hard time telling on which button their finger laid. However, even though results were close, a Friedman test failed to reveal a statistically significant difference between the underlying distributions of the number of recorded errors for interaction method G ($M=0.00$, $SD=0.00$), method C1 ($M=1.57$, $SD=1.90$) and method C2 ($M=0.43$, $SD=0.79$): $\chi^2=5.375$, $p=0.068$, $\alpha=0.05$.

Finally, for each experiment, we also computed the average number of errors per participant across all interaction methods. A Mann-Whitney U test was used to compare the underlying distributions between the two experiments but failed to reveal any statistically significant difference between the two tasks' error rates: $U=13$, $p=0.165$, $\alpha=0.05$.

IX. Discussion

Across all experiments, based on the average times for each interaction method and the smallest absolute bound of the 95% confidence interval for the time differences, it took 14% to 43% less time to perform the tasks using gesture based interactions compared to other interaction methods (C1 or C2). While the results of these preliminary experiments are interesting, our sample size was small (7 participants). A more extensive study with participants drawn randomly from all the population is needed to see if these results could be generalized to the population at large. Moreover, we should point out that we did not counterbalance the order in which participants performed the tasks in the experiments. While unlikely, this could have lead to a fatigue effect and that should be accounted for in future studies.

Our results are nevertheless encouraging as they show that gesture-based interactions open up new possibilities and seem to be more efficient for novice users performing a simple task. Users especially enjoyed brushing away the alerts they did not need. Intuitively incorporating such a feature in a point and click interface is obviously feasible (using for example mouse gestures) but their meaning would still be diluted and the transition between the real and the virtual would still be too prominent.

Gesture-based interactions seem to be a good way to seamlessly query the user about the relevancy / importance of current alerts for novice users. Contrary to other peripherals (i.e. the Twiddler2), no real learning is required as the gestures used are readily understood and even already commonly used by every user. Even for more experimented users, such low attention interaction mechanisms could prove more useful than shortcut keys which become increasingly hard to remember as the number of different avatars and possible actions increase. Our system is relatively scalable (avatars can be displayed as long as a free placeholder is correctly detected) even if it currently relies on artificial markers placed beforehand in the surrounding environment.

X. Future Work

We plan on conducting more experiments, with novice and eventually with more experienced users, to better measure the performance gain (compared to conventional interfaces) and to try to assess the limits of gesture-based interactions.

A. *Other Interaction Mechanisms*

We already experimented with various other interactions mechanisms to broaden the possible actions that could be used. For example, after fitting a RFID reader into the sole of a modified shoe we investigated kicking and stomping actions. For example, "stomping" an avatar could mean that the embodied software process is of no use to the user and the alert/information is to be discarded. "Kicking" an avatar could be used to delay an alert according to the intensity of the kick. Even though results and user feedback were positive, such actions are not applicable in all situations, could be tiring in the long run and could even be socially unacceptable.

We also believe that there is no need to develop a complete and complex gesture-based command language covering every possible interaction with our avatars. We view gesture-based interactions as a low-attention mean to quickly and efficiently manage alerts in a mobile setting. They are best suited to initiate the communication with an avatar, not sustain a deep / long dialogue (other more conventional interaction paradigms will probably be better suited for this once an avatar is selected).

B. *Real World Applications*

To better illustrate the merit of gesture-based interaction with embodied virtual processes, one could imagine managing alerts while driving a car. In such a setting, paying attention to the real world is paramount and a driver simply cannot afford costly transition into the computer interface space to manipulate a pointer (which requires precise hand-eye coordination) or another peripheral while driving. However, using placeholders located for example on the car dashboard, it would be possible to convey meaningful information to the user through virtual avatars and quickly obtain a user feedback while ensuring that the focus of his/her attention stays in the real world.

XI. Conclusion

The key contribution of this paper was to present and test a new interaction paradigm where a user can interact with virtually embodied software processes through tangible and meaningful real-world actions. Our research shows that such interactions are feasible (albeit in a controlled environment) and, for the first stages of the interaction with the software process/agent, seem to lead to a statistically significant reduction in task times (at least 14%) compared to more conventional interaction methods. We did not observe, however, a statistically significant difference in error rates. Nevertheless, a more extensive study is needed to investigate this approach and more work is required to validate and apply this concept in a practical way. In the long run, users will most likely be able to naturally interact with the many software agents competing for their attention and seemingly embedded into reality.

Acknowledgment

The authors would like to thank the reviewers for their insightful and constructive comments.

References

[1] T. Starner, "The Challenges of Wearable Computing: part 1&2," *IEEE Micro*, vol. 21, pp. 44-67, 2001.

[2] T. Starner, "Attention, memory, and wearable interfaces," *IEEE Pervasive Computing*, vol. 1, pp. 88-91, 2002.

[3] W. Piekarski and B. H. Thomas, "ThumbsUp: Integrated Command and Pointer Interactions for Mobile Outdoor Augmented Reality Systems," in *Proceedings of HCI International*. Greece, 2003.

[4] J. Zucco, B. H. Thomas, and K. Grimmer, "Evaluation of Three Wearable Computer Pointing Devices for Selection Tasks," in *Proceedings of the 9th International Symposium on Wearable Computers (ISWC'05)*. Tokyo, Japan: IEEE, 2005, pp. 178-185.

[5] A. Schmidt, "Implicit Human Computer Interaction Through Context," *Personal Technologies*, vol. 4, pp. 191-199, 2000.

[6] G. D. Abowd and E. D. Mynatt, "Charting past, present and future research in ubiquitous computing," in *Human-Computer Interaction in the new millenium*, J. M. Carroll, Ed.: Addison-Wesley, 2002, pp. 511-536.

[7] A. Feldman, E. Munguia Tapia, S. Sadi, P. Maes, and C. Schmandt, "ReachMedia: On-the-move Interaction with Everyday Objects," in *Proceedings of the 9th International Symposium on Wearable Computers (ISWC'05)*. Tokyo, Japan: IEEE, 2005, pp. 52-59.

[8] N. Plouznikoff, A. Plouznikoff, and J.-M. Robert, "Object Augmentation through Ecological Human - Wearable Computer Interactions," in *Proceedings of the 2005 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob2005)*. Montréal, Canada: IEEE, August 2005, pp. 159-164.

[9] B. Rhodes and P. Maes, "Just-in-time information retrieval agents," *IBM Systems Journal*, vol. 39, pp. 685-704, 2000.

[10] T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsh, R. W. Picard, and A. Pentland, "Augmented Reality Through Wearable Computing," *Presence*, vol. 6, pp. 386-398, 1997.

[11] H. Ishii and B. Ullmer, "Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '97)*. Atlanta, U.S.A.: ACM, 1997, pp. 234-241.