Expressive Gesture Controller for an Individual with Quadriplegia

Adam Boulanger
MIT Media Lab, 20 Ames St, Cambridge, MA 02139
adamb@media.mit.edu

ABSTRACT
We present a strategy for mapping and parameterizing expressive gesture for music control with an infrared head-mounted controller. The system allows performers to rely on movement that they find natural despite no prior experience with expressive music performance. It additionally empowers users to identify and contextualize their control parameterization within the key events of a single composition. This highly specialized design strategy is discussed as it relates to our new work in adaptive systems that tailor to the movement and novel contexts provided by ubiquitous users. The technology is piloted with a user with severe motor impairment as a result of quadriplegia. Implications for the field of gesture control and pervasive systems that integrate expressive input are discussed.

Author Keywords
Assistive Technologies, Design, Health, Other applications

ACM Classification Keywords
H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing - methodologies and techniques, H.5.2 [Information Interfaces and Presentation]: User Interfaces - Input devices and strategies, H.5.2 [Information Interfaces and Presentation]: User Interfaces - User-centered design

INTRODUCTION
In the spring of 2005, we brought the Hyperscore composition software to Tewksbury Hospital, a long-term chronic care facility where residents have illnesses as diverse as schizophrenia, cerebral palsy, bipolar disorder, traumatic brain injury, Alzheimer's disease, depression, and other diseases. In the past, we had shown that Hyperscore was an accessible tool allowing novices to create complex and engaging music without any prior experience in the area [6]. At Tewksbury Hospital, the introduction of the technology was evaluated by staff nurse practitioners, occupational and physical therapists as to exhibited patient changes throughout the intervention, functionally, cognitively, and socially.

Throughout the work at Tewskbury, a broader research question emerged regarding our primarily social intervention. If patients were making gains in areas that related to their symptoms, can we distinguish the components of the intervention that are responsible for these changes? Can our technologies that serve as platforms for social change also automatically document the interaction to examine causal change in individuals using the technology?

Secondly, whereas hyperscore provides an interface for any novice user to compose music, we also wanted to develop a specialized interface that would allow a severely disabled individual to control the performance of a piece of music in real-time. Unlike other interfaces for disabled individuals, this technology would address the research question: how to parameterize gesture input for wholly expressive tasks, where users do not have experience with performance or gestures related to expressive performance, but rather, resort to their innate abilities in expressive communication.

The culmination of this three year program was in 2007, where Dan used our technology to perform his music, with expression and control, for over 2000 persons as part of the Media Lab's Human 2.0 event, for over 11,000 people as part of the AARP national conference, and again at the TED conference, in 2008.

RELATED WORK
Despite the performative element of our interface, the developed technology falls within the province of assistive technology, and head-mounted continuous input devices.

Assistive technology in computing is primarily concerned with providing access to existing desktop platform navigation, internet content presentation, text-to-speech and mouse-keyboard emulation for a user-base with sensory functional impairments in addition to the general, typically abled user. The assumption is that the prevading technological platforms can support robust user interaction, if designed defensively for alternative methods of input [4]. Assistive technology design guidelines exist for consideration during the application and content development process [13, 1].

Assistive technologies concerned with user input are typically evaluated with regard to the number of successful entered events in a given period of time, the real-estate occupied by the supporting software on a display space, user-fatigue during continuous control, number of errors in a given time-period, and experimental control observations of alternative input devices [21, 8, 19, 9]. The assumption is that these metrics will serve to iterate the design process and create an interface that is effective, compared to alternatives, for users with diverse functional requirements.

Gesture as an input has been in consideration since the beginning of computer science in the form of character-recognition and sketch interfaces [2, 18]. Text input continues, as well as interfaces supporting more wholistic hand and body motion in addition to applications such as desktop navigation [5], robotic control [3, 10, 22] and gesture-object interfaces [20]. For individuals with motor control impairments, wheelchair control has been driven by both hand gestures [17], and head gesture [23].

Head gesture is also considered in the field of multimodal interfaces [14, 11]. Communication consists of many different streams of information, from speech, to eye-gaze, posture, and body movement. Researchers have proposed leveraging these different streams of input to develop new interfaces. Multimodal input systems are intended to expand accessability to different users and application scenarios other than those supported by traditional desktop interfaces. They often integrate across disparate domains of input [24, 7]. In our application, we are necessarily limited to a small set of possible input modalities, due to the functional limitations of our user. Despite this physical restriction, multimodal systems share similar goals and novel technologies to support input from the domains we are interested in, namely, head movement and gesture analysis.

The broader problem for ubiquity and assistive technology is that current assistive devices are mostly limited to desktop control and text-to-speech applications. As such, our technology is a departure from pre-existing research in assistive technologies, in that it supports a domain of expressive control that is not concerned with desktop computing, or speech production. Our hope is that expressive control can broaden the application of assistive technology to new domains more suitable for pervasive deployment in activities of daily life. To this end, ambient intelligence researchers are beginning to introduce new design principles to bring expressive control within the pervue of pervasive computing [16]. With our application, we offer music performance control, and in conclusion, consider new problems in the adaptation of assistive technology towards ubiquity.

SYSTEM INPUT AND MAPPING

A head mounted infrared point, tracked by color-filtering a live camera feed, allows for several levels of control. Each level presents new assumptions and problems from the standpoint of expressive control. Of primary importance to this project was determining the assumptions of using one control parameter over another, and finding solutions to use control in a way that would not inhibit natural expressive movement. For instance, a raw x-y input gives the user's position in time and a necessary component of expressive movement. However, shape and the change of movement to create shape, could also be important to expressive control. What are the assumptions introduced by measuring certain shapes and not others? If the designers exclusively take shape as a control parameter at the expense of position in time, does this fit with the user's concept of the movements that are essential to the expression that they are attempting to accomplish? These questions invoke a user-centered design process that permeated the construction of the tool, and its parameterization.

The first level of control consists of the mechanics of the movement: velocity, acceleration, and their precedent, position in an x-y plane. At the second level of control we have summary statistics that further describe the mechanics: averages for position, velocity, and acceleration, minimum and maximum range. Already, at the second level, the designers encounter the dilemma of necessarily restricting the type of expressive movement being analyzed as they are forced to choose a bounded period of time to calculate summary statistics from the unbounded, continuous movement exhibited at the first level of control. At the third, and most sophisticated level, there are the models of classification, and categorization that assume the previous analysis and mechanics while fitting them into a model of the what the system designers find to be whole-scale movements within the expressive task. This level also allows for pattern detection.

Expressive performance is not intrinsic in any subset of these three levels of control. After developing the input system, the designer has exposed the wealth of information available for control, but, what does it mean in the terms of a performance? We needed a metaphor to guide the selection of control parameters, their mapping to music characteristics, and the parameterization that would give the mapping variation over time. To find an appropriate metaphor, we examined Dan's movement. Dan is quadraplegic. He only has control of his head movement, and that control is imperfect and limited. His head movements often lack fluidity. It is difficult for him to sustain deliberate and continuous control. Often, his body will go into ataxia, in which involuntary muscle contraction will interrupt his attempt to point with his head or maintain a position with his head. Furthermore, his movement is more pathological in one hemisphere than another. These limitations are critical to the success of head-mounted input devices for disabled individuals [12, 15].

At the beginning of our performance work with Dan, in parallel to our development of the IR input mechanism, we worked with Dan to record his movement both during and after listening to short musical excerpts. Musical excerpts were chosen from various genres and styles, with different numbers of instruments. Musical examples were also chosen with a range of salient expressive features. For instance, in one particular cello example, the performer starts incredibly soft before launching into a frenzied series of notes in the upper register of the instrument. This contrasted less salient exceprts such as the middle of a popular tune, where the expressive characteristics were fairly consistent with little to no departure from a set pattern. The goal was to assess to what extent Dan was able to make concerted, reproducible movements that he felt accurately represented the types of expressive elements that were indicated in the music.

Dan is a naive performer, in the sense that he has no prior experience with movement to affect the resultant interpretation of an artwork. However, he is an incredibly expressive person. His limited head movement, his eye movements, facial expressions, and head gestures do more than express mood, but direct conversation. We had hoped that Dan could bootstrap a similar expressive language for music interpretation. As a result of movement with recordings, we confirmed that Dan was comfortable with whole-scale head movements in time, with some precision. He communicated to us that he was not confident that the types of movements he was making for the presented musical examples were, “correct”. However, we hoped that in the movement surrounding his own compositions, his internal representation of the expressive affordances of his piece would be stronger, and lead to higher confidence in the goodness of fit of his interpretive gestures.

Considering Dan’s movement to music examples, and the physiological limits of his pathological movement, we decided on a metaphor for the mapping based in part on the idea of conducting, and also, partially, on certain aspects of solo instrumental performance. From conducting, we mapped velocity and shape of his curved input line to control orchestration selections. Fast, angular movement changes instruments. Orchestration selections change between sets of instruments during the playback. This selection process is constrained, in the sense that we avoid any type of note selection, in addition to harmony selection or any type of decision process that would significantly restructure the content of the performed work. The application needed to free Dan from any type of algorithmic composing in real-time, to align as closely as possible with Dan’s preconceived notion of the types of control indicative to music performance. A conductor will not surf through a decision tree of possible final compositions. To rely on Dan’s movement as a subset of his learned associations with traditional performance control, instrument selection seemed a sensible mapping target.

In addition to directing changes in orchestration, from solo instrumental performance, we sought to control dynamics over time. Range of motion and shape of the curved input control the amplitude, in time, of the performance. Slower, curved movements allow the user to shape dynamics and create concerted phrases in the performance. The two predominant parameters for expressive control in solo instrumental performance are amplitude and time. Because of dans pathological movement, which is sensitive in the time domain, we decided to filter the movement in time, and to avoid a mapping that would force dan to be a timekeeper, constantly pushing and pulling the tempo of his performance. Dynamics are conducive to large-scale, shaping movements. Time as a parameter requires repetitive and constant feedback to maintain a meter with variation, and is less suitable to control as an abstraction from an x-y input.

MEASUREMENT IN CONTEXT

After the input technology, and the basic mapping were established, we began to work with Dan as we would with any artist. Repeated performance using the system lead us to determine that mapping needed to be tuned relative to the piece and performance that Dan wanted to give. It is not enough to define an input system, and a set of mapping targets born out of a performance metaphor. Despite the fact that a concert pianist has his input system in place, the wealth of his mapping opportunities vary continuously throughout a performance. Expressive parameters are not dealt with equally at all times in a performance. At some points, based on the piece of music, and the plan of the performer, certain parameters dominate the performance. The dynamics and shape of some melodic lines may come to the forefront of the artists attention and delivery. Dynamics may as a whole remain flat, in a moment to play with timbre or the quality of sound. All of these decisions are made in parallel. For our technologies to be truly expressive, they need to be contextualized to the pieces and moments in which they afford an individual to be expressive.

Through repeated rehearsal of a single piece of Dans music, we engaged in a discussion with Dan where we would ask him what he wanted to achieve from one section of the piece to the next. The software includes a score follower, tracking each note event in the performance. This serves to provide us with an index, which we then use to reassign mapping, and to parameterize the scale and sensitivity of the mapped parameters from moment to moment. The result is that the intial mapping targets, dynamics and orchestration, vary in their sensitivity as well as their assignment to various sections of the performance. For instance, a moment of the piece where the first string accompaniment enters, Dans system stops the targeting of orchestration changes and allows for a moment of highly sensitive dynamics control assigned to the incoming string accompaniment. The result is that Dan is able to bring in the accompaniment and taper its dynamics with a high degree of accuracy, before the sensitivity on the orchestration switching increases, in addition to other changes that allow Dan to focus on more whole-scale phrasing, in effect, turning away from the string section which was disproportionately important for that one moment in the piece.

UBIQUITOUS COMPUTING FROM PERSONALIZED INSTRUMENTS

Weve defined this area of research as personalized instruments. With mapping based on Dan’s movement affordances, the gestures that he finds meaningful to deliver a performance, and with the parameterization of the mapping contextualized on a single piece of music, the system is highly tailored for one individual. However, the strongest implications for the work pertain to adapting the input and design principles of the project to expand this technology for anybody.

Our controller required developing, in the hospital, with a set of software tools that would allow us to quickly redistribute mapping and parameter values following iterative rehearsal. The barrier to this work being adopted as a controller with ubiquitous distribution, in the hands of any performer, with differences in experience with expressive music control, and completely separate from the desktop paradigm of current music production and computerized performance systems, is automatic parameterization based on the events that an arbitrary user define as meaningful to their interpretation. This poses a significant new research question in the domain of gesture control, assistive technology, and universal accessability. As computing environments become pervasive to our daily lives, for expressive input to be considered as a method of control design strategies need to investigate not only how input can be adapted to multiple devices, but how the input can be adapted to novel definitions of what constitutes expressive movement. The result of this research was the process in which we were able to determine what movement Dan found meaningful for musical control. Subsequently, we were able to wrap an input technology to fit that movement in the context of the previously composed piece and begin new work in systems that can automatically parameterize from any user's proposed range of expressive movement.

With us leaning over Dan's shoulder, he's able to communicate to us that he needs the mapping to be reassigned or resensitized to one event or another in the piece. Our research is now moving in the direction of the development of an interface that allows users to flag the events for a musical score that that a user finds to be critical to their expressive interpretation - key moments. Then, as a user inputs gestures to modify orchestration and dynamics, the system learns their gestures in the context of musical events. Through successive approximations of characteristic events and their gesture pairs, the system can begin to make assumptions as to the baseline for mapping assignment and parameterization when given a novel score, and a users markings as to the relevant events. To support this future work, and address our interest in documenting the interaction with expressive tools, the movement data used to drive the application can be used not only during this parameterization process, but also to investigate changes in a user's movement over experimental control applications.

Dan is physically limited in his ability to conduct expressive movement. This belies the fact that he is an expressive person. Knowledge of expression without the sufficient structure to allow an individual to apply that knowledge is replete in the general population. To move in the direction where our technologies are prosthetics to enable expressive and creative work from ubiquitous platforms, distributed throughout a population of would-be artists, we must investigate the abstraction of creative processes and how to adapt these abstractions (in the form of mapping and parameter assignments) for as many works and as many individuals as possible.

REFERENCES

1. Web accessibility initiative (wai), http://www.w3.org/wai/.
2. W. Bledsoe and I. Browning. Pattern recognition and reading by machine. In *EJCC '59*.
3. C. Breazeal, G. Hoffman, and A. Lockerd. Teaching and working with robots as a collaboration. In *AAMAS '04*.
4. S. Carter, A. Hurst, J. Mankoff, and J. Li. Dynamically adapting guis to diverse input devices. In *Assets '06*.
5. P. Dhawale, M. Masoodian, and B. Rogers. Bare-hand 3d gesture input to interactive systems. In *CHINZ '06*.
6. M. M. Farbood, E. Pasztor, and K. Jennings. Hyperscore: A graphical sketchpad for novice composers. *IEEE Comput. Graph. Appl.*, 24(1), 2004.
7. M. Hanheide, C. Bauckhage, and G. Sagerer. Combining environmental cues & head gestures to interact with wearable devices. In *ICMI '05*.
8. A. Huckauf and M. H. Urbina. Gazing with peyes: towards a universal input for various applications. In *ETRA '08*.
9. P. Isokoski and R. Raisamo. Quikwriting as a multi-device text entry method. In *NordiCHI '04*.
10. J. Juster and D. Roy. Elvis: situated speech and gesture understanding for a robotic chandelier. In *ICMI '04*.
11. S. Keates and P. Robinson. The use of gestures in multimodal input. In *Assets '98*.
12. E. LoPresti, D. M. Brienza, and J. Angelo. Computer head control software to compensate for neck movement limitations. In *CUU '00*.
13. C. Nicolle and J. Abascal, editors. *Inclusive Design Guidelines for HCI*. Taylor and Francis, London, 2001.
14. Z. Obrenovic and D. Starcevic. Modeling multimodal human-computer interaction. *Computer*, 37(9):65–72, 2004.
15. I. Rauschert. Adaptive multimodal recognition of voluntary and involuntary gestures of people with motor disabilities. In *ICMI '04*.
16. P. Ross and D. V. Keyson. The case of sculpting atmospheres: towards design principles for expressive tangible interaction in control of ambient systems. *Personal Ubiquitous Comput.*, 11(2):69–79, 2007.
17. J. Segen and S. Kumar. Video acquired gesture interfaces for the handicapped. In *MULTIMEDIA '98*.
18. J. Sibert, M. G. Buffa, H. D. Crane, W. Doster, J. Rhyne, and J. R. Ward. Issues limiting the acceptance of user interfaces using gesture input and handwriting character recognition (panel). In *CHI '87*.
19. O. Tuisku, P. Majaranta, P. Isokoski, and K.-J. Räihä. Now dasher! dash away!: longitudinal study of fast text entry by eye gaze. In *ETRA '08*.
20. C. Vaucelle and H. Ishii. Picture this! film assembly using toy gestures. to be published in Ubicomp '08.
21. J. O. Wobbrock, J. Rubinstein, Sawyer, and A. T. Duchowski. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *ETRA '08*.
22. A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka. Precision timing in human-robot interaction: coordination of head movement and utterance. In *CHI '08*.
23. I. Yoda, K. Sakaue, and T. Inoue. Development of head gesture interface for electric wheelchair. In *i-CREATe '07*.
24. T. Yonezawa, N. Suzuki, S. Abe, K. Mase, and K. Kogure. Cross-modal coordination of expressive strength between voice and gesture for personified media. In *ICMI '06*.