# GESTURAL COMMUNICATION WITH ACCELEROMETER-BASED INPUT DEVICES AND TACTILE DISPLAYS

Paul D. Varcholik*
Institute for Simulation and Training
University of Central Florida
Orlando, FL 32826

James L. Merlo, Ph.D.
LTC, US Army
US Military Academy
West Point, NY 10996

## ABSTRACT

In this work, we introduce a communication system for common military hand and arm gestures which does not require a visual connection between the transmitter and receivers. Specifically, we present a computer-mediated gesture recognition system that employs a wireless, accelerometer-based input device for collecting and classifying one- and two-hand and arm gestures. This system delivers message output through an audible channel and through a tactile display. The tactile display emulates the hand and arm signal's spatial qualities through a sequence of vibrations delivered via an elastic belt worn around the soldier's waist. Initial results show promise for this novel way to communicate with multimodal messages to augment visual messaging in challenging and stressful environments where visual messaging may not always be possible.

## 1. INTRODUCTION

According to an Army field manual (FM 21-60) on visual signals, "efficient combat operations depend on clear, accurate and secure communication among [personnel]," (Department of the Army, 1987). When vocal means of communication are inadequate, which is often on the noisy battlefield, visual signals can be an effective alternative for transmitting orders, information, or requests for aid or support (Merlo, Szalma, & Hancock, 2007). However, these signals require line-of-sight between the transmitter and receiver. There are numerous situations where a direct visual connection is unavailable and there is compromise on the ability to exchange critical information through conventional communication pathways. Nighttime operations, inclement weather, man-made and natural terrain obstructions, or concealment often impede visual communication attempts. To overcome some of these issues, "daisy-chaining" or relaying a message is frequently used to send a message through a group. However, this method can delay the transmission from the original sender to the intended recipients. Moreover, visual communication demands a focus on the visual modality possibly distracting a receiving soldier's visual attention from alternate, perhaps more urgent, tasks.

In this paper, we introduce a communication system for common military hand and arm signals which does not require a visual connection between the transmitter and receivers. Moreover, the system instantaneously delivers the message to all recipients and can do so through a variety of output devices. Input gestures, those intended for transmission, are accepted through a wireless accelerometer-based input device and classified by a machine learning algorithm for subsequent delivery. Of particular importance for gesture delivery is the incorporation of a tactile display, an output device which emulates the gesture's spatial qualities through a sequence of vibrations delivered via an elastic belt worn around the soldier's waist. These tactile sequences can be made to intuitively mimic the corresponding hand signal, and research has shown that tactile communication can succeed in conveying information to a recipient even under high physiological stress (Merlo, Stafford, Gilson, Hancock, 2006).

This communication system leverages the extensive amount of hand and arm signal training currently

provided to military personnel. A transmitting user need not learn a new set of commands in order to employ this platform. Instead, the required hand signals are encoded into the gesture recognition system's machine-learning algorithm, providing user-independent and/or user-specific gesture examples. A message recipient does require a training period, when using the tactile display, in order to learn which tactile sequences correspond to which hand and arm signals. However, research has shown that the length of this training period can be mitigated if the tactile signals are intuitively mapped from their visual-spatial equivalent (Gilson, Redden, & Elliot, 2007).

This paper introduces a novel communication system, and provides detail on initial experiments and applications. Also presented is the software application used in training the machine learning algorithm and in teaching a recipient on mapping hand and arm signals to tactile sequences. Finally, we discuss avenues for moving this technology from the lab into the field. Specifically, we look at embedding accelerometers into a soldier's gloves and porting the gesture recognition and transmission systems to pocket-size computing devices.

## 2. RELATED WORK

This work combines two technology challenges, one involving the input of a communication and one the output. Previous studies on communication delivery (output) have shown that tactile systems can produce relatively stable performance improvements across a variety of body orientations even when spatial translation is required (Oron-Gilad, Downs, Gilson, and Hancock, 2007; Terrence, Brill, & Gilson, 2005) and in the presence of physiological stress (Merlo, Stafford, Gilson, & Hancock, 2006). Additional work on tactile studies can be found in (Gilson, Redden, & Elliot, 2007; Prewett, Yang, Stilson, Gray, Coovert, & Burke, 2006).

The system also incorporates audible delivery as an alternative to, or reinforcement of, the tactile output system. Most of human information processing uses multiple sensory inputs, such as the synthesis of visual and auditory cues (Hancock, 2005; Spence & Driver, 2004; Stein & Meredith, 1993). Literature on experiments that involve the use of two modalities of information presented redundantly each show improvement in the areas of accuracy and response time (Spence & Walton, 2005; Gray & Tan, 2002; Strybel & Vatakis, 2004).

For processing input, the system centers on a computer-mediated gesture recognition system. Considerable work has been done in this area, particularly since the early 1990s. The *Glove-Talk* system from Fels and Hinton (1993) used an instrumented glove and machine learning system to recognize a 203 gesture-to-word vocabulary. They improved their work in 1996 and 1998 with the *Glove-Talk II* system (Fels & Hinton, 1996, 1998). In addition to glove-based gesture recognition systems, much work has been done using computer-vision techniques. For example, Binh, Shuichi, and Ejima (2005) developed a vision-based recognition system using Hidden Markov Models (HMM, a machine learning technique).

Finally, Kratz, Smith, and Lee (2007) created a recognition system, similar to the one being presented, which uses the Nintendo® Wiimote and HMM. Our system differs in the choice of machine learning algorithm and through the extraction of a feature set to describe a 3-Dimensional gesture created using the Wiimote.

## 3. GESTURE RECOGNITION SYSTEM

The gesture recognition system is made of three primary components: an input device to collect the movements of the gesture; a software system to classify those movements; and an output device to communicate the gesture to a recipient.

### 3.1 Input Device

The system employs an accelerometer-based input device for collecting gesture data. Specifically, the Nintendo® Wii Remote Controller (Wiimote, see Figure 1) was chosen as an inexpensive, commercial-off-the-shelf (COTS) motion controller with 3-axes of input. Although initially designed for use on the Nintendo® Wii, the Wiimote has been adapted for use on the personal computer (PC). The Wiimote connects to a PC wirelessly, using the Bluetooth communication protocol. Data can be collected from up to four Wiimotes simultaneously.

The Wiimote sends accelerometer data at a maximum frequency of 100Hz. The incoming X, Y, Z floating-point values, from the accelerometers, are treated as "point" information to represent a position in 3D space. In fact, this data is not a true 3D position, and merely indicates forces applied to the accelerometers through

corresponding motion. These values cannot be made to accurately represent a 3D position without an additional sensor (e.g. a multi-axis gyro). However, the data still yields accurate gesture recognition by treating the data as 3D positions (drawbacks of this approach are discussed in section 4). As such, the movement of the Wiimote can be visualized as a connected series of points contained within a 3D bounding volume created by the extents of the point set. For each sample, the 3D point data is augmented with timestamp information to allow for the calculation of speed and distance between samples.

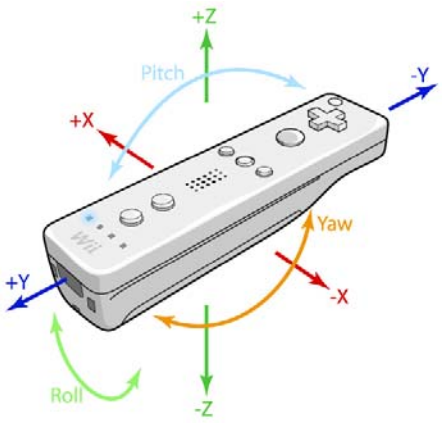


**Figure 1. The Wiimote controller (Troillard, 2008)**

### 3.2 Software

The software system is organized into two distinct components: an underlying library for defining the data structures and machine learning algorithms for gesture recognition; and a graphical user interface for demonstrating the functionality of the system.

At the heart of the underlying gesture recognition library are three data structures: *WiimotePoint*, *Gesture*, and *TrainedGesture*. The *WiimotePoint* object contains the raw Wiimote X, Y, Z accelerometer values and associated sample timestamp. A collection of *WiimotePoints* are used to construct a *Gesture* object – which is an unlabeled representation of a complete one-handed gesture. It is a *Gesture* object that is passed to a machine learning algorithm for classification (labeling). *Gesture* objects expose a set of 29 features extracted from the set of contained *WiimotePoint* objects which are used as inputs into a machine learning algorithm for training and subsequent classification. These features were derived from Rubine's work on 2D symbol recognition (1991) and are listed in Table 1. These features are a key component of the software system, and differentiate this element of the work from that of Kratz et al. (2007).

**Table 1. Feature set**

| Feature | Description |
|---|---|
| Duration | Total duration of the gesture. |
| Max X, Y Z | The maximum x, y, and z accelerometer values (3 features). |
| Min X, Y, Z | The minimum x, y, and z accelerometer values (3 features). |
| Mean X, Y, Z | The mean x, y, and z accelerometer values (3 features). |
| Median X, Y, Z | The median x, y, and z accelerometer values (3 features). |
| Bounding Volume Length | The length of the diagonal of the bounding box created from the extents of the x, y, and z values. |
| Starting Angle | The angle created between the first and third samples points (3 features measuring the sine and cosine of the angle within the XY plane and sine within the XZ plane). |
| First-Last Angle | The angle created between the first and last samples points (3 features measuring the sine and cosine of the angle within the XY plane and sine within the XZ plane). |
| Total Angle Traversed | Summation of the angles between each pair of points (2 features measuring the XY and XZ planes). |
| Total Angle Traversed Absolute | Summation of the absolute values of the angles between each pair of points (2 features measuring the XY and XZ planes). |
| Total Squared Angle Traversed | Summation of the squared values of angles between each pair of points (2 features measuring the XY and XZ planes). |
| First-Last Point Distance | The distance between the starting and ending points. |
| Total Gesture Distance | Summation of the distance between each pair of points. |
| Max Acceleration Squared | The squared value of the maximum acceleration detected between each pair of points. |

A *TrainedGesture* object is a labeled collection of *Gesture* objects, where the gestures are data samples used for training a machine learning algorithm. When recognizing an unlabeled gesture, the set of known

*TrainedGesture* objects is the source for classification. All of these data structures – *WiimotePoint*, *Gesture*, and *TrainedGesture* – are serializable for creating reusable sets of trained gesture data. Importantly, no calculated data (e.g. the features extracted from the *Gesture* objects) are included in the serialization. This allows modifications to such calculation within the software library without invalidating previously created data sets.

Machine learning algorithms complete the gesture recognition library, and are the elements that perform the actual recognition. The library specifies a software interface for implementing various algorithms. In this fashion the library allows for multiple algorithms to be implemented and used against the same type of data. To date, we have experimented with a linear classifier, AdaBoost, and an Artificial Neural Network.

The second software component, the graphical user interface, is presented in Figure 2.

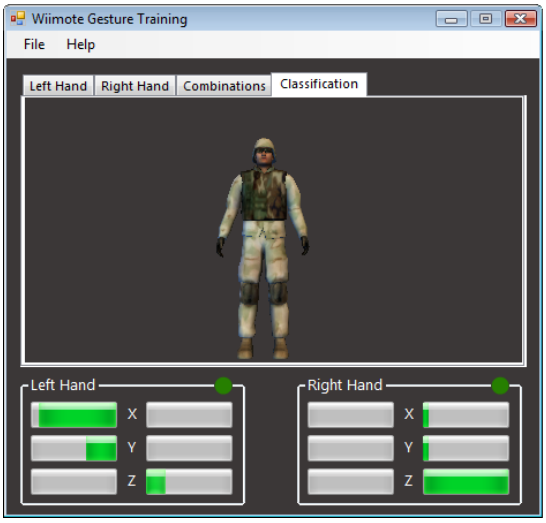


**Figure 2. Gesture training user interface**

This component is used to collect gesture samples, train personnel on recognizing tactile display output, and for validating the performance of the recognition system through visual feedback of the classification process. At the bottom of the user interface, there are status indicators on which Wiimotes are connected and the X, Y, Z accelerometer data being read from them. The Left Hand, Right Hand, and Combinations tabs are used for training the gestures that will subsequently be used for classification. Combination gestures (both left and right hand) can be trained with both hands simultaneously. Previously trained and serialized gesture sets can be loaded through the File menu.

Classification tests are conducted through the corresponding tab which presents an animated 3D model of a soldier that mimics the classified gesture. Audible and tactile feedback is also generated during classification. The animations, sounds, and tactile sequences are pre-programmed and mapped to the trained gestures. Any newly trained gestures that are not mapped to an animation, sound, or tactile sequence, simply present the name (label) of the gesture upon classification. When training or classifying a gesture, the user indicates the start and end of the gesture by pressing the Wiimote's trigger button (B button). This is the same technique for both right and left handed gestures (combination gestures require pressing both left- and right-hand trigger buttons).

The graphical user interface need not be the software system employed by soldiers during actual transmission and receipt of a visual signal. Indeed, for dismounted infantry requiring lightweight and easily transported technologies, a more appropriate system would run on a mobile computing platform and would not provide visual feedback as with the animated 3D model. A simple network protocol has been developed for transmitting the gestures over UDP/IP. The transmitting end provides audible feedback of the recognized gesture being transmitted while the receiving end allows for both audible and tactile delivery of the gesture.

Although initially developed on a personal computer running Microsoft Windows, the software described in this section could be adapted to operate on mobile platforms and over alternate electronic communication protocols.

### 3.3 Ouput Device

There are two output devices supported by the system: audible speakers and a tactile display. Audible speakers relay a sound file mapped to the corresponding gesture. This sound file need not simply be the label of the gesture, and could be used to communicate a more complex message.

For conditions where audio messages are not appropriate, the system incorporates a tactile display. The vibrotactile actuators (tactors) in this system are manufactured by Engineering Acoustics, Inc. They are

essentially acoustic transducers that displace 200-300Hz sinusoidal vibrations onto the skin. Their 17gm mass, when properly loaded on the skin, is sufficient for activating the skin's tactile receptors. The tactor's contactor is 7mm, with a 1mm gap separating it from the tactor's aluminum housing. The tactor is a tuned device, meaning it operates well only within a very restricted frequency range, in this case approximately 250Hz. The tactile display itself is an elastic belt like device with eight tactors attached (see Figure 3). The belt is made of elastic and high quality cloth similar to the material used by professional cyclists. When stretched around the body and fastened, the wearer has an actuator over the umbilicus and one centered over his or her spine in the back. The other six actuators are equally spaced, three on each side, for a total of eight (Cholewiak, Brill, & Schwab, 2004).

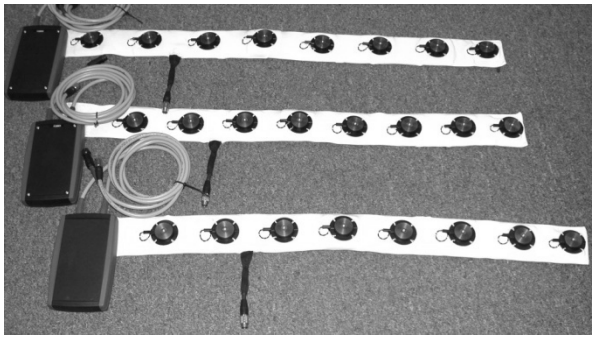

**Figure 3. Three tactile display belt assemblies**

The tactors are operated using a Tactor Control Unit (TCU) which is a computer-controlled driver/amplifier system that switches each tactor on and off as required. This device is shown on the left side of the tactile display belts in Figure 3. The TCU weighs 1.2lbs independent of its power source and is approximately one inch thick. This device connects to a power source with one cable, to the display belt with the other, and uses Bluetooth technology to communicate with the computer driven interface. Tactile messages were created for the chosen gesture set (discussed in section 4) where the vibrations approximate the patterns presented with the visual signal.

## 4. DISCUSSION

We have experimented with this system using a number of gesture sets and the machine learning software system has demonstrated a high level of accuracy. In an informal user study with 5 participants, we experimented using 7 gestures from the Army field manual (FM 21-60) on visual signals: Attention, Disregard, Halt, Increase Speed, Mount Up, Start Engines, and Stop Engines. The results of this study show a 96% classification accuracy when supplying 30 training samples per gesture to the linear classifier. The system had accuracy above 94% with as few as 10 samples per gesture. These results are comparable to those found in Kratz' HMM implementation where a maximum 95% accuracy was reported (Kratz et al. 2007). Future work is required to verify and formalize these results, and to make recommendations on machine learning algorithms and training sizes for optimal results.

Apart from the technical details of the system described in the previous sections, the focus of this communication platform is on delivering visual signals without a visual connection between transmitter and recipients. In internal demonstrations the system has operated as described – exhibiting its capability in recognizing pre-trained gestures and delivering both audible and tactile output to a recipient. Moreover, we have experimented with delivering gestures to multiple recipients simultaneously. This has likewise been very successful, but there are outstanding research questions on techniques for selecting individuals for specific delivery. At present, delivery of a message is broadcast to all recipients connected to the same electronic network (e.g. UDP broadcast). However, we have yet to exercise the full system in the field to determine its efficacy at communicating between individuals or groups. Note that the tactile display has been studied extensively in its ability to successfully convey information, but not as part of the complete communication system presented here (see Gilson, Redden, & Elliot, 2007).

For many applications, the Wiimote is not the right choice for a deployment of this system. It is used here as a proof-of-concept chosen for its low cost, and availability. We believe our system is adaptable for any 3-axis accelerometer-based input device or even to a more capable device (e.g. a device with augmenting gyroscopes) which could provide genuine position data. With the Wiimote, there are ambiguities in the data that affect the recognition accuracy of gestures – particularly for gestures that are similar. We suspect that our system has an upper-limit to the number of gestures that can be accurately classified. Further study is required to determine what that upper-bound is, but a more capable input device would likely provide better performance. Additionally, differentiating similar static poses is not possible with the Wiimote. For example, the data sampled

from the Wiimote, when held at shoulder height, cannot be distinguished from the same orientation held above the users head. The Wiimote must be in motion, and those motions must be somewhat different for each gesture, in order to accurately train a machine learning algorithm.

A final note on the Wiimote is on the choice to use the trigger button as the gesture start/end indicator. An alternate technique is to continuously sample the data stream looking for recognized gestures. The linear classifier requires approximately 10ms to recognize a gesture using our 7 gesture data set at 30 samples per gesture. With such high performance, it is feasible to continuously query the data stream, but it is likely that this will recognize gestures prematurely. For example, the system is capable of distinguishing a half circle, from a full circle. If the classifier continuously queries the data stream, it might classify the gesture as a half circle before the user completed the gesture. There are ways to mitigate this result and further study is recommended to determine the most appropriate technique for starting/ending a gesture. This is a significant consideration for input devices that may not include a button-style triggering mechanism. An accelerometer-instrumented glove, for example, would need a non-button-style triggering technique to prevent unintended transmissions and accurate recognition. Pinch gloves and pre- and post-gestures are two potential alternatives for triggering gesture start/end.

The question of learning complex tactile communication signals, especially for use in adverse or unusual circumstances is an important future issue. The tactile system performs a redundancy gain as the receiver of the signal now has an additional means of receiving communication if the user can hear the signal as well. Initial testing seems to result in superior performance for tactile communication and traditional hand and arm signals combined. Stimulus response compatibility will have to be analyzed carefully to maximize performance as we consider different ways to input and relay visual signals. However, when individuals are faced with extreme challenges and the traditional sources of information are either diminished or eliminated altogether, the system described in this paper provides an important alternative communication channel and one that may be exploited.

## ACKNOLWEDGEMENTS

This work is supported, in part by the Office of Naval Research and also by the US Army Research Laboratory under Cooperative Agreement N00014-07-1-0098. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR, ARL or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

The authors of this paper would like to acknowledge the efforts of Chris Ellis at the University of Central Florida for his work in the development of the gesture recognition system and associated machine learning feature set.

## REFERENCES

Binh, N. D., Shuichi, E., & Ejima, T. (2005). Real-time hand tracking and gesture recognition system. *GVIP 05 Conference*, Cairo, Egypt.

Cholewiak, R. W., Brill, J. C., & Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. *Perception and Psychophysics, 66*, 970–987.

Department of the Army. (1987). Visual Signals. (Field Manual No. 21-60). Washington, DC: Government Printing Office.

Fels, S. & Hinton, G. (1993). Glove-talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks, 4(1), 2-8*.

Fels, S. & Hinton, G. (1996). Neural networks for computer-human interfaces: Glove-TalkII. *Proceedings of the International Conference on Neural Information Processing (ICONIP96)*, Hong Kong, 1299-1304.

Fels, S. & Hinton, G. (1998). Glove-TalkII: A neural network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks, 9(1), 205-212*.

Gilson, R. D., Redden, E. S., & Elliot, L. R. (Eds.). (2007). Remote Tactile Displays for Future Soldiers. (Vol. ARL-SR-0152): Aberdeen Proving Ground, MD 21005-5425.

Gray, R., & Tan, H. Z. (2002). Dynamic and predictive links between touch and vision. *Experimental Brain Research, 145*(1), 50-55.

Hancock, P.A. (2005). Time and the privileged observer. *KronoScope, 5(2)*, 177-191.

Kratz, L., Smith, M., & Lee, F. (2007). Wiizards: 3D Gesture recognition for game play input. *Proceedings of the 2007 conference on Future Play*, Toronto, Canada.

Merlo, J.L., Stafford S.C., Gilson, R. & Hancock, P.A. (2006). The effects of physiological stress on tactile communication. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, CA.

Merlo, J., Szalma, M., & Hancock, P.A. (2007). Stress and performance: Some Experiences from Iraq. In: P.A. Hancock and J.L. Szalma (Eds.). *Performance under stress*. Ashgate: Aldershot, England.

Merlo, J.L., Terrence, P.I., Stafford, S., Gilson, R., Hancock, P.A., Redden, E.S., Krausman, A., Carstens, C.B., Pettit, R., & White, T.L. (2006). Communicating through the use of vibrotactile displays for dismounted and mounted soldiers. *Proceedings of the 25th Annual Army Science Conference*, Orlando, FL

Oron-Gilad, T., Downs, J.L., Gilson, R.D., & Hancock, P.A. (2007). Vibro-tactile cues for target acquisition. *IEEE Transactions on Systems Man, and Cybernetics: Part C, Applications and Reviews, 27(5)*, 993-1004.

Prewett, M. S., Yang, L., Stilson, F. R. B., Gray, A. A., Coovert, M. D., Burke, J. (2006). The benefits of multimodal information: A meta-analysis comparing visual and visual-tactile feedback. *Proceedings of the 8th International Conference on Multimodal Interfaces*, 333-338.

Rubine, D. (1991), Specifying gestures by example. *Proceedings of the 18th annual conference on Computer graphics and interactive techniques, 329-337.*

Spence, C., & Driver, J. (Eds.). (2004). Crossmodal Space and Crossmodal Attention. Oxford; New York: Oxford University Press.

Spence, C., & Walton, M. (2005). On the inability to ignore touch when responding to vision in the crossmodal congruency task. *Acta Psychologica, 118*(1), 47-70.

Stein, B. E., & Meredith, M. A. (1993). The merging of the senses. Cambridge, MA: MIT Press.

Strybel, T. Z., & Vatakis, A. (2004). A comparison of auditory and visual apparent motion presented individually and with crossmodal moving distractors. *Perception, 33*(9), 1033-1048.

Terrence, P. I., Brill, J. C., & Gilson, R. D. (2005). Body Orientation and the Perception of Spatial Auditory and Tactile Cues. *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, Orlando, FL.

Troillard, C. (2008). Wiimote Image. http://www.osculator.net/wiki/ uploads/Main/pry-wiimote.gif. Accessed Sept. 2008.