

“Will Use It, Because I Want to Look Cool”

A Comparative Study of Simple Computer Interactions Using Touchscreen and In-Air Hand Gestures

Vidya Vaidyanathan and Daniel Rosenberg

San Jose State University
vidya.vn@gmail.com, dan@rcdoux.com

Abstract. The Xbox Kinect and now the Leap Motion Controller have brought about a paradigm shift in the way we interact with computers by making the recognition of 3D gestures affordable. Interfaces now understand natural user interfaces, integrating gestures, voice and various other kinds of multimodal input simultaneously. In this paper we attempted to understand in-air gesturing better. The purpose of the study was to understand differences between touchscreen and in-air gesturing for simple human computer interactions. The comparison of the gestures was done in terms of Muscle effort/fatigue and Frustration, Satisfaction and Enjoyment We have also tried to study the learnability of in-air gesturing. In our research we found that in-air gesturing was significantly superior with respect to muscle effort and fatigue when compared with touchscreens. We also found that in-air gesturing was found to be more fun and preferred because of its “coolness factor”. Lastly, in-air gesturing had a rapid learning curve.

Keywords: HCI, Touch Screens, in-air gestures, ergonomics, EMG, learnability, social acceptability, natural user interfaces (NUI).

1 Introduction

Gone are the days when user interfaces were based entirely on buttons, joysticks, keyboards and mice. Today the world has advanced into direct manipulation devices such as touchscreens and smart phones. An external device that maps onto the x-y coordinate system of a computer control is no longer required. The future of the computing world lies in interfaces described in the press as gesture controlled, motion-controlled, direct, controller-less and natural. The most popular gesture controlled devices that exist in the market today are gaming devices such as the Nintendo Wii, the Microsoft Kinect, the Sony Eye Toy and the Leap Motion. Smart phones and tablets are joining the trend of using gestures.

1.1 What Is Gesture Recognition?

Gesture recognition mainly concerns with identifying, recognizing and making meaning of human movements. The human body parts involved can be the hands, arms,

face, head or the body [1]. Kendon states that amongst all human body parts conveying gestures, the hand gestures are the most natural and universal [7]. They form a direct and instant form of communication. Hand gestures are therefore the most used method for interaction with technological systems [7]. According to Dr. Harrison at Carnegie Mellon University, the human hands alone are capable of tens of thousands of gestures, individually and in combination. Some tasks hinder the use of hands to interact with devices, such as checking email while driving a car [2]. Atia et al showed that in such cases certain applications use face and body related gestures. They also showed that using the leg to gesture was limited due to the spatial constraints [2].

1.2 Background and Related Work

In a survey, where Americans were polled for the top two inventions that improved their quality of life, “television remote” and “microwave oven”, emerged as the winners [17]. Freeman and Weissman explored the control of a television using gesture recognition [17]. They compared voice and gesture as two candidates for equipment control. Voice had the advantage of having an established vocabulary, but was deemed not appropriate for the context. Gestural control was more appropriate for the context, but lacked a natural vocabulary [17].

Perzanowski et al explored the possibility of building a multi-modal interface based on voice and gestures [13]. Their interface used natural gestures especially those made using the arms and hands. They made use of meaning-bearing gestures that were associated with locational cues for a human-robot interaction. The meaning-bearing gestures mainly included indication of distances (by holding the hands apart) or directions (tracing a line in the air). When a user says “go there”, the accompanying gesture signaling the direction was essential to make sense of the verbal command. Perzanowski et al observed that in noisy environments, gestures was largely used to compensate for lack of comprehensible auditory input [13].

The idea of using “free hand” gestures as an input medium is based out on the famous “put that there” experiment conducted in 1979. This experiment used primitive gestural input in the form of gestural languages: Task control primarily used gestures. Sign language interpretation was one of them. Some other examples were those where Sturman [16] presented a gestural command system to orient construction cranes, while Morita et al showed the use of gestural commands in an orchestra [12].

1.3 Naturalness of Gestures

The more natural a gesture is to its context and the more coherent in its mapping to human performance, the higher its interaction fidelity will be [4]. Bowman et al conducted a series of experiments to answer the questions they posed. They found that increased “interaction fidelity” has an increasingly positive experience on the user performance and efficiency of user tasks. Natural gestures were especially beneficial

when the tasks were more complex. Users perceived that interactions with a higher degree of interaction fidelity were more fun, engaging and had higher immersive value.

Considering learnability of a NUI, Wigdor and Wixon claim that a NUI is one that provides a quick and enjoyable learning experience from novices to skilled users [18]. This rapid learnability occurs due to practice. They also define an NUI to be extremely enjoyable.

1.4 Social Acceptability of Gestures

Beyond recognizability, the acceptability of gestures is also critical. Certain cultures have politeness conventions for gestural use [8]. For example, pointing with the left hand is considered impolite in the country of Ghana. Here, receiving and giving with the left hand is also considered taboo [9, 10]. Hand gestures might have some drawbacks, such as acceptance or rejection in a public space [2]. Atia et al found that public found it threatening when a user performed the gesture of a large circle in a public place [2]. Studies have examined the usability of hand gestures in different generic environments, especially public places [15]. Ronkienen et al conducted “tap gesture” based experiments where they presented participants with gesture-based scenarios and quizzed on their willingness to use the gesture in various situations [15]. It was observed that the social acceptability of performing a gesture was dependent on where it is performed and the audience it was performed for. Further, certain gestures could be viewed as threatening in public spaces. Rico and Brewster expanded Ronkienen’s experiment and examined the social acceptability of eight common gestures, example wrist rotation, foot tapping, nose tapping, shoulder tapping, etc. [14]. They showed that acceptability depends on the combination of audience and workplace. For example in the US, nose tapping was acceptable when the performer was alone at home, or in a pub among strangers, but not when alone in a workplace or in front of friends and family.

1.5 Drawback of Gestures

Baudel and Beaudouin-Lafon extensively explored the limitations of any gesture recognition system [3]. Fatigue was found to be one the key limitations. Gestural communication used more muscular activity than simple keyboard interaction, mouse interaction or speech. The wrist, fingers, hands and arms all contributed to the commands. In order for the gestures to be of minimal effort, they had to be concise and fast. Over time they may induce fatigue in the user [5]. Among the more recognized tools to measure muscle fatigue is the Electromyographic (EMG) analysis [5]. The surface EMG has limitations related to electrode placement, skin impedance and cross-talk [11]. In spite of the limitations, the surface EMG has been shown to be a valid and reliable tool to identify muscle fatigue [5].

2 Experimental Design

The purpose of our study was to understand differences between touchscreen and in-air gesturing for simple computer interactions. Gestures were used to select from a series of tiles displayed on a computer screen. The comparison of the gestures was done in terms of measuring

- Muscle fatigue/effort
- Frustration, satisfaction and enjoyment
- Learnability of in-air gesturing, as a measure of the time component was also measured

2.1 Hypothesis

We hypothesized that in-air gesturing would be preferred to a touchscreen for interacting with a computer and that users would easily learn to use in-air gesturing during the experimental period

2.2 Participants

Thirty-two participants (SJSU) students taking the course Psych 1 and a few volunteers) were recruited to perform the tasks for this study. The participant pool was coordinated with the SJSU Psychology Department. The mean age of the participants was 20 years old and ranged from 18-29 years old. Fourteen participants were male and eighteen were female. Recruitment of the participants was entirely voluntary and scheduling was done online using SONA (human-subject pool management software).

Participants with active musculoskeletal disorders were excluded. This information was elicited by asking the participant about any disorders. All participants, except two were right handed. These two were ambidextrous and conducted the experiment using their right hand. All participants had used a smart phone or tablet with a touchscreen for at least one month.

The study was approved by the SJSU Institutional Research Board (IRB). A consent form, a photo consent form and an NDA was signed by each participant before beginning the testing session.

2.3 Apparatus and Instrumentation

A Dell AIO with an 3rd generation Intel Core i7-3770S processor 3.10 GHz with Turbo Boost 2.0 up to 3.90 GHz configured with 8GB Dual Channel DDR3 SDRAM at 1600MHz was used to conduct the study. Its' display was a 27.0" diagonal wide-screen native resolution (FHD) with tilt base and a Touchscreen with HD support.

The system ran software that emulated the Windows 8 64 bit (Metro) home screen in English. Surface EMG sensors and software provided by Biometrics Ltd. (<http://www.biometricsltd.com>) was utilized. A range of surface EMG pre amplifiers

was used with either the Biometrics DataLink DLK900 or DataLOG P3X8 for monitoring, storing and analyzing muscle electrical activity.

2.4 Procedure

Learnability Section: Task 1. Learnability section: Task 1 All participants completed a learnability task. This task helped familiarize the participant with the equipment (interfaces) and the gestures used to perform the task. This task helped determine if the gestures were easy to learn and remember.

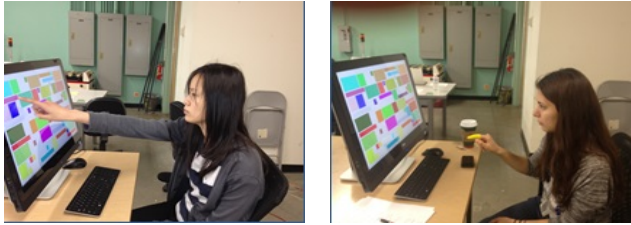


Fig. 1. Tile selection using touch screen and in-air gesturing

The participant was seated upright (back firmly against backrest) on a comfortable chair with armrests. Armrest height, seat height and distance from the screen were adjusted so to be consistent relative to each participant's body size and reach. In preparation for the in-air gesturing, the participant wore a yellow tag on his right index finger. This helped the recognition algorithm detect the finger for in-air gesturing more robustly.

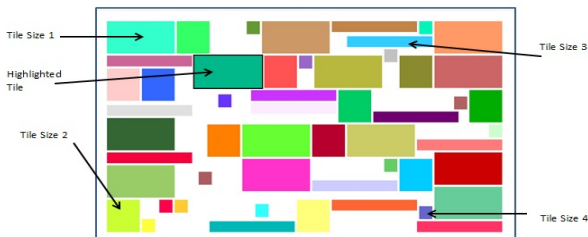


Fig. 2. Input screen with varying tile sizes

In this task, tiles lighted up in a pre-determined order every three seconds and participants selected the highlighted tile. The screen consisted of a collage of “Metropolitan” like tiles of four different sizes. The sizes were 310 x 150 pixels - rectangle shaped tile, 150 x 150 pixels - square shaped tile, 390 x 150 pixels - rectangle shaped tile and 60 x 60 pixels - square shaped tile. The first two sizes were the native Windows 8 desktop icons. The third size was from an email client, a highly used application. The last size was one of the smaller sized tiles prevalently used in Windows 8.

Tiles were separated uniformly by a 10 pixel gutter. Tiles were highlighted in a pre-determined fashion every three seconds. The colors were randomized.

Tiles of any size would highlight by showing a black blinking border around the tile. This task was repeated for both the touchscreen and in-air interface. Participants tapped the screen in a touchscreen interface and moved a finger in free space for in-air gesturing to perform a “selection gesture”. The selection gesture was a “Hold to Click” gesture, where the pointer controlled by the finger was held motionless for about 1.5 seconds on a tile to indicate selection. Less than the 1.5 second hold would result in unsuccessful selection of the tile. On selection of a tile, a graphic was displayed on the software to provide selection feedback. The tile remained highlighted until successful selection of the tile was complete, after which the next tile was highlighted. The hand moved from the resting position (which is the position where the participant is comfortably seated, with no hand lifted up) to the relevant point of selection. The participants selected a total of 20 tiles during the task. The software running the task measured the following factor:

Duration: Response time from the point of tile highlight to selection. Questionnaires were administered to elicit subjective data about the experience for the touchscreen and in-air gesturing interface.

EMG Setup. The surface EMG transducers were placed parallel to the muscle fiber at three locations on the dominant side of the body as in Figure 3. They are the Upper trapezius, Anterior deltoid and Extensor Digitorum (the center of the dominant posterior forearm at approximately 30% of the distance from the elbow to the wrist). The muscle was palpated to detect the exact point of muscle activity when the participant extended his fingers. The ground electrode was connected to the left ankle.

Maximal Voluntary Electrical (MVE) activation measurements were performed against manual resistance to normalize the EMG signals from each location.

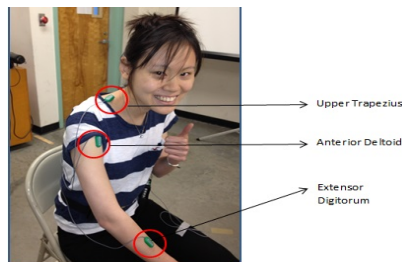


Fig. 3. EMG Transducers fixed to the 3 positions in the body

Task 2. Next, the participant performed Task 2. The task and setting was similar to task 1 but with a different tile layout. Here tiles of a particular size alone were highlighted each time. The task was performed four times, once for each size. Each task took approximately 1-2 minutes. EMG data was recorded for each task. The order of the tasks was randomized to reduce order effects.

Task 3. Next, Task 1 of the experiment was repeated. The duration was measured and compared with the initial session. The comparison helped us understand to what extent learning happened. The same questionnaires were administered once again and later analyzed for any change in subjective measures. Subjective ratings on discomfort and ease of use of the touchscreen and in-air gesturing interface was elicited by means of self-report questionnaires. This was done at the end of task 1 and 3. The questionnaires consisted of check boxes, semantic differential scales and open ended questions. The semantic differential scales used 7 points ranging from very high to very low with a center point of neutral stance.

3 Analysis of Data

3.1 Learnability Task

The time taken to perform the Learnability Task 1 and Learnability Task 2 for in-air gesturing alone were compared. Of 32 participants, 26 showed an improvement in speed in the second session. That made up about 81.25% of the participants. A paired samples T-test was conducted to compare the mean differences between the times taken for the two learnability sessions for in-air gesturing alone. The mean time for Learnability 1 was 142.40 seconds while the mean time for Learnability 2 was only 128.36 seconds. The mean difference was found to be statistically significant, $M=14.04$, $SD=16.82$, $t(31)=4.722$, $p<0.05$, Cohen's $d=0.83$. This shows a large effect in the mean difference. When 1.5 seconds of hold to click time and three seconds between highlights (82.5 seconds) was reduced from each participant's time, we found a statistically significant result with the same t and p values.

3.2 EMG Setup

The participants were subjected to four trials with each one of the four tile sizes. There were eight tiles in each category in all of the permutations. For each participant, six values were obtained which were the average value for Upper Trapezius, Anterior Deltoid and Extensor Digitorium for touchscreen and in-air gesturing.

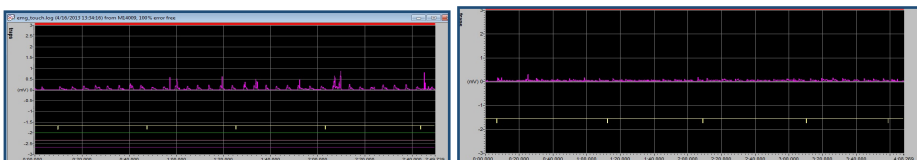


Fig. 4. Filtered Signals for Upper Trapezius for touch screen and in-air gesturing

Figure 4 shows the values for participant 18's upper trapezius values for a touchscreen and in-air gesturing after application of filters. The spikes show activity in the upper trapezius as it moved to select a tile. We barely see any activity in the second graph, showing that in-air gesturing requires less effort when it comes to the upper trapezius. Similarly we saw barely see any activity showing that in-air gesturing requires less effort for the anterior deltoid. The spikes in the touchscreen were

attributed to every time the participant stretches out his arm to touch the screen. For the Extensor Digitorum, some activity was seen with the in-air gesturing when compared to touch screen. The spikes in the touchscreen were attributed to every time the participant closes the wrist to point to the screen.

Three Repeated measure ANOVAs were conducted to compare the muscle effort of the Upper Trapezius, Anterior deltoid and Extensor Digitorum immaterial of the tile size. Significant results were obtained for Upper Trapezius and Anterior deltoid. Very small significance was obtained for Extensor Digitorum.

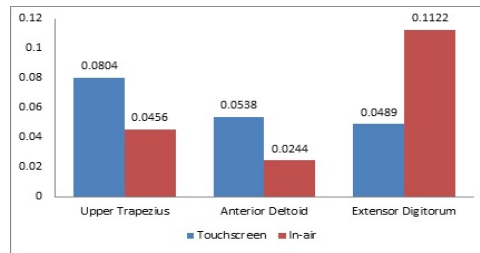


Fig. 5. Comparing means values for the 3 muscle points for touchscreen and in-air gesturing

Further ANOVA analysis showed no statistically significant difference between the four tile sizes for touchscreens. There was statistical significance between the four tile sizes for in-air gesturing. It was found that tile size 4 was the most difficult to manipulate in in-air gesturing.

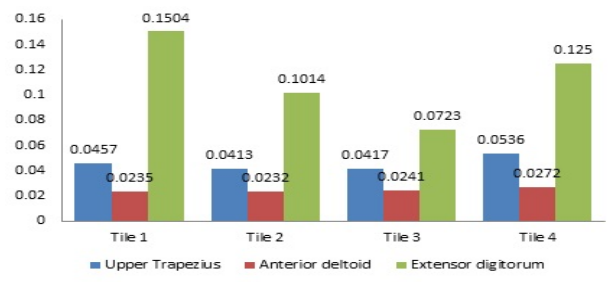


Fig. 6. Comparison of mean values for Tile sizes versus Muscle point for In-air gesturing

3.3 Subjective Questionnaire Analysis

Familiarity with In-Air Gesturing: To begin with, 10 out of 32 participants were familiar with in-air gesturing either through Xbox Kinect etc. while 22 were unfamiliar. About 69% of the participants were unfamiliar with in-air gesturing

Interface Preference: Of the 30 participants, 27 preferred touchscreen at the end of both learnability sessions. Three participants preferred in-air gesturing to begin with. Two participants changed their preference from touchscreen to in-air gesturing. One participant changed preference from in air to touchscreen. One participant’s preference with respect to in-air gesturing remained the same.

Analysis of Individual Questions: Touchscreen data reported represents data recorded after the first learnability session. In-air data represents data from both learnability sessions. All data reported is an average of the individual ratings given by the 32 participants.

Figure 7 compares the values for 5 factors. The scale defined 1=low and 7=high. Touchscreen generally reported the best value. In-air gesturing after the second learnability session reported better values than the first.

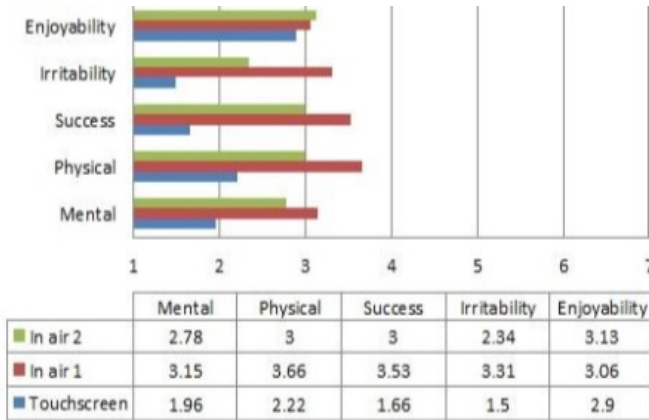


Fig. 7. Comparing above 5 questions for touchscreen, in- air gesturing session 1 and in-air gesturing session 2

No participant felt silly or embarrassed to use the touchscreen. Nine participants felt so using in-air gesturing after the first session. The number fell to four after the second session. For in-air gesturing, the degree of embarrassment was 4 (around mean) after the first session, but fell to a low 2.75 after the second session.

Participants found in-air gesturing initially easier in the first session at a value of 2.84 than in the second session of in-air gesturing, with a value of 3.13

All 32 participants said they would use the touchscreen in a public place. 9 said no to in-air gesturing after the first session, which came down to 7 after the second session. 2 participants changed their preference to yes. Among the 9 participants in the first session, 6 found it silly to use in-air gesturing. 3 were ready to use it in a public place even though they found it silly. After the second session, only 2 out of the 7 found it silly to use in-air gesturing. The number increased to 5 for those people who found it silly but still would use it in a public place.

After the first session, 5 participants didn't want to own a device, while after the second session, the number increased to 2. 5 participants who found in-air gesturing silly, wanted to own a device after the first session. After the second session, 3 who found it silly wanted to own a device. Further, 2 people who found it silly and didn't want to use in-air gesturing in a public place, still wanted to own a device. The various reasons people wanted and didn't want to own a device were multifold. The number of reasons to own a device outnumbered the ones that were against owning a device as seen in Table 1.

Table 1. Reasons quoted verbatim

I want a device	I don't want a device
It's Cool	It will make me self-conscious
Fun and exciting	Too unreliable
For development purposes	Touchscreen is more efficient
Support new technology	Hard to use
Enjoyable for gaming	Hard to control
Easy to use	In accurate
Makes life more efficient	Feel no need for gesturing
Fun and less work for the shoulder	
People get more exercise while using the Kinect kind of things	
For curiosity	
I will eventually get used to it	
Use when hands are not free	
I will use it to create my own gestures	
Fun to do something at a distance than up close	
Don't want to do the extra work in touch-screens	
Only in situations where physical touch is not possible	

4 Discussion

The primary goal of the study was to elicit preference between two interfaces, the touchscreen and in-air gesturing. A secondary goal of the study was to understand the learnability of in-air gesturing as it is a new and upcoming technology, especially given its limitations.

EMG recordings very clearly showed that in-air gesturing was a more ergonomic methodology of interacting with the computer when compared to touchscreen. Statistically significant results were found for two of the critical muscle points, the Upper Trapezius and Anterior deltoid. EMG recordings also showed that muscle effort increased significantly when the size of the target decreased. Among the 4 tile sizes uses, tile size 4 was found most difficult to select during in-air gesturing and differed statistically significantly from the other 3 sizes.

The experiment session lasted for about 1 hour 15 minutes, approximating about 1 hour of time between the first learnability and second learnability sessions. That accounts for a total of 8 minutes maximum of in-air gesturing. It was found that there was a statistically significant improvement in the time taken to perform the two similar sessions. Mean value of the time taken decreased by 14.04 seconds. This was found to be a large effect. Note that 69% of the participants were using in-air gesturing for the first time in their lives during the experiment. This shows that significant learnability happened over a period in in-air gesturing within an hour of time.

Subjective questionnaire analysis showed a similar trend throughout. Touchscreen always rated better for almost all the questions over in-air gesturing. But between the two sessions of in-air gesturing, ratings after the second session were always found better than the first. It is evident that with time and practice, in-air gesturing is comparable to touchscreen eventually for almost all the factors. The only factor that saw a higher rating in the second session was the “ease of use” of in-air gesturing.

Majority of participants did not find it silly to use in-air gesturing and were ready to use it in a public place. There were some conflicting answers such as, some participants who found it silly, were ready to use it in a public place. Some participants who found it silly and were not ready to use it in a public place still wanted to own a device. Majority of participants wanted to own a device capable of in-air gesturing. The most popular reason was because they found it cool, among various other relevant reasons.

According to Harrison’s definition, the gesture does not directly indicate its intent because in real life, “hold to click” does not indicate selection [6]. It is interesting to note that though this gesture is not very intuitive for selection purposes, the learning curve was found to be very easy and showed statistical significance. This goes against the literature that claims that it is the naturalness and intuitiveness of a gesture that defines the learning curve.

We learnt in this experiment that “social acceptability” does play a role. But this experiment has also shown that this self-consciousness of people actually fades away with time and people would want to use a device because in-air gesturing is considered more technologically advanced. It overrides the social taboo of in-air gesturing.

5 Conclusion

In-air gesturing definitely emerged as a winner during the period of the experiment. Participants clearly showed that it was easy to learn the technique of interacting with the interface and the subjective attributes were comparable to that of the current reigning touchscreen with the passage of time. The reasons why participants preferred in-air gesturing outnumbered the reasons why participants preferred the touchscreen and the reasons were variant and spread across a large spectrum. In-air gesturing emerged as the winner ergonomically when compared to touchscreen. This experiment has shown what Steve Jobs once quoted that touchscreen computers are “ergonomically terrible”. It has also been shown that participants prefer the in-air gesturing to touchscreen because of the coolness factor associated with new technology.

Acknowledgements. My special thanks to Dr. Louis Freund and Dan Rosenberg of San Jose State University whose guidance remained from start to end of the experiment. My thanks also go to the employees of the startup who made possible the experiment. I would like to acknowledge Joyce Kii and all the participants who made valuable contributions to the research. I would also like to thank my family who have been my stronghold through this journey.

References

1. Acharya, T., Mitra, S.: Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 37(3), 311–324 (2007)
2. Atia, A., Takahashi, S., Tanaka, J.: Smart gesture sticker: Smart hand gestures profiles for daily objects interaction. In: *Proceedings of: 9th IEEE/ACIS International Conference on Computer and Information Science, IEEE/ACIS ICIS 2010, Yamagata, Japan, August 18-20 (2010)*
3. Baudel, T., Beaudouin-Lafon, M.: Charade: Remote control of objects using free-hand gestures. *Communications of the ACM – Special Issue on Computer Augmented Environments, Back to the Real World* 36(7) (1993)
4. Bowman, D., McMahan, R., Ragan, E.: Questioning naturalism in 3D user interfaces. *Communications of the ACM* 55(9), 78–88 (2012)
5. Christova, P., Kossev, A., Kristev, I., Chichov, V.: Surface EMG recorded by branched electrodes during sustained muscle activity. *J. Electromyogr Kinesiol* 9, 263–276 (1999)
6. Harrison, C.: Meaningful gestures,
<http://www.economist.com/node/21548486> (retrieved March 3, 2012)
7. Kendon, A.: *Gesture: Visible action as utterance*, pp. 326–355. Cambridge University Press, Cambridge (2004)
8. Kita, S.: Theoretical issues in nonverbal behaviors. Presentation Slides (2007), retrieved from <http://ling75.arts.ubc.ca/cogs//cogs401>
9. Kita, S., Essegbey, J.: Pointing left in Ghana: How a taboo on the use of left hand influences gestural practice. *Gesture* 1(1), 73–95 (2001)
10. Kita, S., Danzinger, E., Stolz, C.: *Cultural Specificity of Spatial Schemas manifested in spontaneous gestures*. MIT Press, Cambridge (2001)
11. McQuade, K., Dawson, J., Smidt, G.: Scapulothoracic muscle fatigue associated with alterations in scapulohumeral rhythm kinematics during maximum resistive shoulder elevation. *JOSPT* 28(2), 74–80 (1998)
12. Morita, H., Hashimoto, S., Ohteru, S.: A Computer Music System that Follows a Human Conductor. *IEEE Computer*, 44–53 (July 1991)
13. Perzanowski, D., Schultz, A., Adams, W., Marsh, E., Bugajska, M.: *Building a multimodal Human-Robot interface* (2001)
14. Rico, J., Brewster, S.: Usable gestures for mobile interfaces: Evaluating social acceptability. In: *Proceedings of CHI 2010*, pp. 887–896 (2010)
15. Ronkainen, S., Hakkila, J., Kaleva, S., Colley, A., Linjama, J.: Tap input as an embedded interaction method for mobile devices. In: *Proceedings of TEI 2007*, pp. 263–270. ACM Press (2007)
16. Sturman, D.: *Whole-hand input*, Ph.D thesis, Media Arts & Sciences. MIT Press (1992)
17. Weissman, C., Freeman, W.: Television control by hand gestures. In: *IEEE Intl Workshop on Automatic Face and Gesture Recognition* (June 1994)
18. Wigdor, D., Wixon, D.: *Brave NUI World*, 1st edn. Morgan Kaufmann, Burlington (2011)