

Thumbs Up to Gesture Controls? A Cross-Cultural Study on Spontaneous Gestures

Alison Meier¹, Kelly Goto¹, and Michael Wörmann²

¹ gotomedia & UX Fellows, 2169 Folsom St. m-302, San Francisco, CA 94110, USA
{alison,kelly}@gotomedia.com

² Facit Digital & UX Fellows, Neuhauser Strasse 17, D-80331 Munich, Germany
m.woermann@facit-digital.com

Abstract. As a growing number of three-dimensional motion and gesture-controlled consumer electronics come to market, there are new challenges for user experience design for globally merchandized devices, because gestures naturally vary between regions and cultures. The UX Fellows network undertook a study to learn how people from different countries would gesture to control consumer electronics. Eighteen agencies interviewed 360 people from 18 countries around the world to see how people would spontaneously gesture to achieve specific TV commands. Results from each country were aggregated and compared to determine which countries used the same gestures and which used different ones.

Keywords: gesture, motion, culture, international, consumer electronics.

1 Introduction

Commercial products that respond to three-dimensional gestures are becoming accessible in the market, first in video game consoles like Xbox Kinect, then in smartphones like the Samsung Galaxy S4, and now in computers via the Leap Motion device. Despite the commercial availability of devices controlled by gestures and even their low price point, no common gesture language exists. As technology becomes increasingly capable of understanding that we are signaling with our hands, the practicality of communicating in this manner relies on the design of easy to use gesture languages.

Gestures are usually defined as interactions where hands move in the air in a way that replaces a more traditional mouse and pointer or touchscreen interaction. Some gestures function as pointing gestures, where the hand or finger simply replaces the pointer as used in a mouse. These are called pointing gestures. This study examines a different kind of gesture, what we call semantic gestures, where the hand (or hand and arm) performs series of movements that are recognized by the device. We define a gesture language as a set of semantic gestures.

The goals of a gesture language are twofold: first, the machine must be able to recognize and interpret distinct gestures. Second, the gestures must incur minimal burden

on the user by being both easy to use and easy to remember [1]. “Easy to use” includes the need for gestures to require minimal physical exertion as well as be socially acceptable [2].

A previous study [3] found that user-defined gestures—in which individuals choose their own gestures for each action—are easier to remember than pre-defined gestures. Yet user-defined gesture languages would lead to infinitely fragmented languages, in which users would have to be willing to and capable of creating their own set of gestures for each device. An alternative to user-defined gestures is “user-elicited” gestures, in which “sets are generated by a representative group of users in a study...and are then carefully compiled and selected to form an ‘agreed upon’ gesture set that is still recognizable by the system” [3, 4].

As a user-elicited gesture language should be defined by a “representative group of users,” the global nature of consumer electronics markets indicates that the representative group should consider cross-cultural expectations regarding gestures. Although some previous work [5] acknowledges cultural differences in gestures, no previous study has attempted to understand the specific instances of cultural disagreement over three-dimensional gestures. A past study [6] of touchscreen gestures found high agreement across cultures for direct manipulation gestures and low agreement for symbolic gestures. Three-dimensional gestures are inherently more symbolic than touchscreen gestures, so presumably the disagreement would increase as well.

2 The Study

UX Fellows, a network of user experience research agencies around the world, undertook a study to understand how people from around the world would interact with consumer electronics using gestures, in particular the television. Participants were asked to imagine they could control a TV using semantic gestures and to demonstrate how they would perform a series of commands.

Participants then rated how difficult they found it to determine each gesture. Eighteen agencies (each representing a different country) participated in the study. Each conducted identical sessions with 20 participants in its home country for a total of 360 participants around the world. The following countries were represented:

- | | |
|-------------------|-------------------------|
| 1. Argentina (AR) | 10. Italy (IT) |
| 2. Australia (AU) | 11. South Korea (KR) |
| 3. Belgium (BE) | 12. Mexico (MX) |
| 4. Canada (CA) | 13. Netherlands (NL) |
| 5. China (CN) | 14. Russia (RU) |
| 6. Germany (DE) | 15. Spain (ES) |
| 7. Finland (FI) | 16. Turkey (TR) |
| 8. France (FR) | 17. United Kingdom (UK) |
| 9. India (IN) | 18. USA (US) |

2.1 Recruiting Methodology

Each agency recruited participants in its home country using the same criteria. The criteria included the following:

- Regular use of an HD flat screen TV (manufactured within the past 6 years)
- Regular use of at least three consumer electronics devices or services such as a Blu-Ray player, video on demand, home theater system, or gaming consoles
- Regular use of a touchscreen device such as a smartphone, tablet, or MP3 player
- No past use of gesture-controlled devices such as Xbox Kinect or a gesture-enabled TV
- Electronics-savvy consumers (a high interest in electronics and on top of the latest technologies)
- Likely to buy consumer electronics from major international brands (no national, retailer, or “no name” brands)

Pre-tests indicated that usage of touchscreen devices influenced gestures selected, so UX Fellows opted to bias toward touchscreen users rather than away from them, as more and more people are being exposed to touchscreens over time.

Participant ranged from ages 16 to 65, with a mean age of 32. Forty-nine percent of the participants were female.

2.2 Session Methodology

Sessions were conducted on an individual basis with one participant at a time with a moderator. Each session lasted about 15 minutes.

Care was taken to set up the session environment in a consistent manner by each agency. Participants were placed in front of a turned-off flat screen TV. A moderator sat to the side of the TV, while participants sat on a couch or comfortable chair.

As a warm-up exercise, participants were asked to demonstrate everyday gestures used in real life (rather than with electronics) such as “I can’t hear you,” “call me,” and “very tasty.” Then the participant was asked to rate how easy it was to think of a gesture (1 = easy to 5 = difficult). After the warm-up, participants were asked to demonstrate how they would gesture to the switched off TV to complete a series of commands. For example, the moderator would ask, “How would you switch on the TV?” and the participant would demonstrate a gesture and then explain why they selected that gesture. They were asked to rate the how easy it was to think of the gesture the same as with everyday gestures.

They were told in advance that their gestures did not need to be a cohesive set (for example, volume up did not need to be the inverse of volume down). The moderator was careful not to use gestures while speaking, to avoid biasing the responses. The session was captured by video recording.

3 Results

Results of each country's sessions were aggregated and then compared to the other countries. The TV tasks were grouped in three categories—high consistency, medium consistency, and low consistency—based on the number of countries that reported the majority of participants using the same gesture for the same command. On average, TV task gestures had a higher average mean difficulty (2.1) compared to everyday gestures (1.6) (1 = easy to 5 = difficult).

3.1 Everyday Gestures

There was high consistency between most everyday gestures. Gestures that were identical across most countries include “be quiet,” “call me,” and “write an email.” Signaling numbers 2, 4, and 5 were nearly identical between countries, though there were differences when demonstrating the number 3: a few countries used the thumb and first 2 fingers, while most used the first 3 fingers.

A few gestures were similar for the majority of countries except for a couple of countries. For example, “come closer” was different in East Asian countries than most countries. Chinese participants gestured “I don't know” by waving the hand with the palm facing forward, while participants from other countries shrugged shoulders.

Most countries used a thumb pointed up to gesture “OK,” although a significant minority had the index finger meet the thumb in a circle with the other fingers spread. “Very tasty” and “I can't hear you” varied greatly between countries.

Everyday gestures were generally easy to think of and had an average mean difficulty of 1.6. The most difficult everyday gesture to think of was “write an email” with a mean difficulty of 3.0. Despite this difficulty, almost all countries used the same gesture to demonstrate this concept.

3.2 TV Tasks: High Consistency Gestures

Some gestures had a high degree of consistency between countries with all or all but a few countries using the same gesture (see Table 1). Most of these gestures represented basic commands, such as changing the volume, fast-forwarding or rewinding, or pausing video-on-demand (VOD). Participants rated these gestures as generally easy to think of, with rating averages from 1.5 to 2.0.

Table 1. Gestures with a high degree of consistency between countries

Command	Most Frequent Gesture	Mean Difficulty	Countries with divergent gestures
Volume up	Lift hand or index finger	1.8	0
Volume down	Downward movement of hand, palm down	1.5	0
Pause current program	Fingers spread palm forward	2.0	1 (NL)

Table 1. (continued)

Pause VOD	Fingers spread palm forward	1.5	1 (NL)
Confirm on-screen dialog (OK)	Thumb up	1.9	3 (CN, IT, KR)
Fast forward VOD	Arm wave right to left	1.9	3 (AU, NL, UK)
Rewind VOD	Arm wave left to right	1.7	3 (AU, NL, UK)

3.3 TV Tasks: Medium Consistency Gestures

Another group of gestures were consistent for about half or 2/3 of the countries (see Table 2). Gestures that had a medium level of consistency between countries included slightly more complex commands such as volume up and down, muting sound, and cancel. The mean difficulty of these gestures increased slightly to an average range of 1.6 to 2.3. In this category, a number of participants used gestures borrowed from touchscreen interactions, such as swiping to navigate a menu. The direction that they swiped (left to right or right to left) varied between countries.

Table 2. Gestures with a medium degree of consistency between countries

Command	Most Frequent Gesture	Mean Difficulty	Countries with divergent gestures
Channel up	Hand swipe right to left	1.8	5 (AU, CA, ES, KR, US)
Stop VOD	Fingers spread palm forward	2.1	5 (CN, FI, KR, NL, RU)
Channel down	Hand swipe left to right	1.6	6 (AU, CA, ES, KR, UK, US)
Mute volume	Index finger to lips	2.0	7 (AU, BE, CN, FI, MX, RU, UK)
Backward navigation in menu	Hand swipe left to right	2.3	8 (AR, AU, BE, CA, ES, MX, RU, US)
Deny on-screen dialog (cancel)	Thumb down	1.8	9 (AR, AU, BE, CN, IT, KR, MX, RU, TR)
Skip chapter VOD	Draw a bow from left to right	2.2	9 (BE, CN, ES, FI, FR, IN, MX, NL, UK)

3.4 TV Tasks: Low Consistency Gestures

The final cluster of gestures had very low agreement between countries, with only few countries agreeing on a gesture. Difficulty averages also increased, ranging from 2.0 to 3.2. These gestures were primarily complex interactions with minimal real-world counterparts, such as share current program and context for current program. On the other hand, extremely basic commands such as switching on and off the device also fell in this category.

Table 3. Gestures with a low degree of consistency between countries

Command	Most Frequent Gesture	Mean Difficulty	Countries with divergent gestures
Switch on device	Fist with thumb pressing down (like remote control)	2.0	10 (BE, CA, CN, DE, FI, FR, IT, KR TR, US)
Previous chapter VOD	Draw bow from right to left in air	2.0	10 (BE, CN, ES, FI, FR, IN, MX, NL, UK, US)
Switch off device	Clap hands	2.9	10 (AR, CA, ES, IN, IT, KR, MX, NL, RU, TR)
Restart VOD	Palm facing TV	2.0	11 (AR, AU, BE, DE, ES, KR, MX, NL, RU, UK, US)
Share current program	Spread both hands upward to both sides	3.2	12 (AR, CN, FI, FR, IN, KR, MX, NL, RU, TR, UK, US)
Electronic program guide	Both arms at chest swung out wide	2.8	13 (AU, BE, CN, FI, IN, IT, KR, MX, NL, RU, TR, UK, US)
Context info for current program	Spread hands	2.8	14 (AR, AU, BE, CN, DE, ES, FI, FR, IN, KR, RU, TR, UK, US)
Home page	Arm in circular motion	2.9	14 (AR, AU, BE, CA, DE, ES, FI, IN, IT, MX, NL, TR, UK, US)

3.5 TV Tasks: Lack of Clustering among Regions

We found it notable that there was often little consistency between neighboring countries, even when the countries could be considered culturally similar. For example, Canada and the United States disagreed on gestures over a quarter of the time. The Netherlands and Germany disagreed over half the time (participants from the Netherlands frequently used gestures that were different from the global majority.)

On the other hand, some gestures were identical in countries that are in very different regions and have minimal cultural similarity. For example, participants from Australia, China, and Mexico all used the same gesture to indicate mute volume (moving hand from right top to left bottom). This gesture was different from the majority gesture.

3.6 Desirability of Gestures to Control Consumer Electronics

In open-ended discussion following demonstrating gestures, participants generally reported interest in controlling their electronics with gestures, in particular the “coolness factor” of using a new technology and not being tied to a remote control.

Participants did express concerns about using gestures, however—in particular needing to learn a gesture “language,” the accuracy with which the machine recognizes the gestures, interference from other people in the room or unintentional gesturing, and being comfortable while watching TV.

3.7 Study Limitations

Many countries are not represented in this study, including entire regions such as Africa. In addition, each agency conducted the research in the city in which it is located, so many regions within each country are not included. For example, the United States agency conducted the research in San Francisco, California, which may not be representative for the country as a whole. Ideally, future versions of this study will include more countries and larger samples within each country.

In addition, participants in each country often used different gestures. Results were considered based on the majority usage. In many cases, a gesture may have been used by a minority of participants in the country that may be worth noting, especially if it was used by participants in other countries, either in a minority or majority capacity.

4 Conclusion

Findings support the possibility of creating a global gesture language for most basic TV interactions. With the exception of turning the TV on and off, most typical TV interactions could be completed with semantic gestures that people in most countries find natural.

For complex tasks, it is possible to use different gestures on a country-by-country basis, but because we did not observe regional clusters, this could lead to a highly fragmented gesture language that varies greatly from location to location.

As technology continues to advance and devices become even more capable of accurately recognizing semantic gestures, the success of gesture interactions will depend on the gestures themselves. Creating a gesture language for complex tasks for a global market that is easy to remember, socially acceptable, and easy to perform will be a significant challenge for designers and developers.

Acknowledgments. Thank you to all of the UX Fellows agencies who participated in this study and contributed to analyzing and reporting the results. Full results as well as a video of participants performing some of the gestures can be found at <http://www.uxfellows.com/gesture.php>.

References

1. Oh, U., Findlater, L.: The challenges and potential of end-user gesture customization. In: Proc. CHI 2013, pp. 1129–1138 (2013)
2. Ashbrook, D., Starner, T.: MAGIC: a motion gesture design tool. In: Proc. CHI 2010, pp. 2159–2168 (2010)
3. Nacenta, M., Kamber, Y., Qiang, Y., Kristensson, P.: Memorability of pre-designed and user-defined gesture sets. In: Proc. CHI 2013, pp. 1099–1108 (2013)
4. Wobbrock, J., Morris, M., Wilson, A.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092 (2009)
5. Vatavu, R.: User-defined gestures for free-hand TV control. In: Proc. of the 10th European Conference on Interactive Tv and Video, pp. 45–48 (2012)
6. Mauney, D., Howarth, J., Wirtanen, A., Capra, M.: Cultural similarities and differences in user-defined gestures for touchscreen user interfaces. In: Ext. Abstracts CHI 2010, pp. 4015–4020 (2010)