Perspectives of Gestures for Gestural-Based Interaction Systems

Towards Natural Interaction

Nur Zuraifah Syazrah Othman[a], Mohd Shafry Mohd Rahim[a], Masitah Ghazali[a] & Sule Anjomshoae[b]
[a]Universiti Teknologi Malaysia, [b]Umeå University

A frequently mentioned benefit of gesture-based input to computing systems is that it provides naturalness in interaction. However, it is not uncommon to find gesture sets consisting of arbitrary (hand) formations with illogically-mapped functions. This defeat the purpose of using gestures as a means to facilitate natural interaction. The root of the issue seems to stem from a separation between what is deemed as gesture in the computing field and what is deemed as gesture linguistically. To find a common ground, this paper explores the fundamental aspects of gestures in the literature of psycho-linguistic-based studies and HCI-based studies. The discussion focuses on the connection between the two perspectives – in the definition aspect through the concept of meaning and context, and in the classification aspect through the mapping of tasks (manipulative or communicative) to gesture functions (ergotic, epistemic or semiotic). By highlighting how these two perspectives interrelate, this paper provides a basis for research works that intend to propose gestures as the interaction modality for interactive systems.

Keywords: gesture, natural interaction, gesture-based computing, human–computer interaction (HCI)

Othman, Nur Zuraifah Syazrah, Mohd Shafry Mohd Rahim, Masitah Ghazali & Sule Anjomshoae "Perspectives of Gestures for Gestural-Based Interaction Systems: Towards Natural Interaction". *HUMAN* ***IT*** 14.3 (2019): 1–25. © The authors. Published by University of Borås.

Various forms of gestures have been proposed for computing applications. These gestures carry with them the information needed to control a system, and are collectively called a set, dictionary or vocabulary of gestures. Despite the apparent success of gesture-based systems, designing a suitable set of gestures remains a challenging task.

From the seminal work of Bolt (1980) to the more recent works in gesture-based computing (e.g., Dingler *et al.* 2018; Rusnák *et al.* 2018), researchers have continued to propose new forms of gestures. It is not uncommon to find gesture sets designed based on easy recognition purposes. However, it is also often that such gesture sets consist of arbitrary hand formations with illogical mappings (e.g., holding three fingers means "walking"). It is quite disconcerting that gestural interfaces which have been touted as "natural" do not resonate at all with the natural bodily movements that they are supposed to represents. Perhaps the race to take advantage of the advances in technology drives the perceived notion that a gesture interface is universally the best interface for any given application, causing designers to aim for a more "generic" interface instead of the most effective (Nielsen *et al.* 2003; Norman 2010; Norman & Nielsen 2010).

As computers are, even now, increasingly used to facilitate human communication, the early proposal of van Dam (1997; 2001) to consider the knowledge of how humans communicate with each other in the creation process of gestures still stands. Humans mainly communicate using languages, and studies on how humans acquire, use, comprehend and produce languages are mostly from the field of psycholinguistics. But communications are also done through gesturing (Abner, Cooperrider & Goldin-Meadow 2015), and this is where gesture and language are often thought of as a single system (McNeill 2006). Studies in gesture-based computing especially have shown the importance of understanding the collaboration process of verbal and gestural interactions (Evans, Wobbrock & Davies 2016).

Kendon (1986) and McNeill (2006), whose works are influential in the gesture field, were also known to draw the study of gestures through the approach of psycholinguistics. Other works have also shown the tendency to include reference from psycholinguistic studies when designing gestures (e.g., Billinghurst & Vu 2015; Kistler & Andre 2013; Medrano, Pfeiffer & Kray 2017; Rateau, Grisoni & Araujo 2014; Silpasuwanchai & Ren 2015; Tscharn *et al.* 2017; Wobbrock, Morris & Wilson 2009). Indeed, from our review, gestural types common in psycholinguistics are broadly adopted in gestural-based human computer interaction (HCI) research. However, in spite of this trend, much of the gesture sets produced are disparate to human's natural movements (e.g., Angelini *et al.* 2014; Choi *et al.* 2012; Heimonen *et al.* 2013; Urakami 2014; Wu & Wang 2012).

To explore more on this disparity issue, we propose going back to the roots of the gesture domain, as we believe that a clear understanding of definitions and classifications should be the basis in any gesture-based interaction model. As such, we have conducted a review of established definitions and classifications of gestures from both psycholinguistic-based studies and HCI-based studies. As highlighted earlier, the perspective from psycholinguistic studies was also included to reflect the trend of the literature that includes this perspective in HCI-based studies.

Based on our review, we specifically chose to discuss the definitions from the view proposed by Kendon (1986) and Kurtenbach and Hulteen (1990) as we found that both works have been highly influential in shaping the fundamental theories in the gestural literature. Similarly, for classifications of gestures, views from Cadoz and Wanderley (2000), Quek (2003) and Karam and Schraefel (2005) were discussed as we found them to be the more prominent views being cited in the literature.

First, a 'compare and contrast' approach was taken to explore the differences and similarities of the two perspectives. This serves as the basis for the discussion where we highlight how the two perspectives connect in the definition and classification aspect of gestures. We believe that this understanding helps bridge the gap between theory and application of gesture-based interaction, and further provides a shared foundation between human-to-human interaction, human–environment interaction and human–computer interaction.

Furthermore, we feel that the interrelation between the two perspectives is an important factor to consider when deriving gestures for computing systems, illustrated in the conceptual framework presented at the end of the discussion. It is hoped that the framework could facilitate researchers and developers in the designing of suitable gestures gearing towards natural interaction.

Exploring the differences

In this section, we first present the definition of gesture from psycholinguistic and HCI/computing perspectives proposed by Kendon (1986) and Kurtenbach and Hulteen (1990) respectively. Following this, the discussion on classification of gestures is presented. For the psycholinguistic view the discussion is based on the functional classifications of epistemic, ergotic and semiotic gestures proposed by Cadoz and Wanderly (2000), while for the HCI/computing view the discussion is based on classifications of gestural-based interactions of manipulative, semaphoric and conversational gesture systems proposed by Quek (2003).

Defining Gesture

In the matter of defining gesture, Kendon (1986) proposed looking from the mental aspect of a participant in an interaction. In an interaction, a 'gesture' is a behavior that is deliberately expressive and treated as intentionally communicative. The term deliberately expressive denotes movements that share certain dynamic characteristics that "stood out" from the view of the participants in the interaction. Such movements will be regarded by observers as fully intentional (deliberate) and intentionally communicative.

Kendon (1986) also claimed that actions that were done when nervous, such as hair patting or nail biting, though expressive, even during an interaction is neither deliberately done nor intentionally communicative. These gestures are often treated as habitual or involuntary. However, Kendon (1986) pointed out that a practical action, like throwing a waste paper into a basket can also be done in a highly expressive manner, for example, by mimicking a three-point basketball shot. Such deliberate expressive behavior can turn a theoretically meaningless action into a performance of sort. Hence it is observed as intentionally communicative, and thus becoming a gesture.

Another interesting point from Kendon (1986) is that, if practical actions can be given some of the qualities of gesture, it is also possible that a gesture may sometimes be disguised so that it no longer appears as such. What may be a gesture in one situation may 'appear' to be incidental in another. There have been several instances of derogatory gestures being 'camouflaged' or 'suppressed', turning it into incidental mannerisms, such as scratching one's cheek using only the middle finger. The objective of the sender is to not make it seem like it is directed to anybody; that it is essentially harmless and non-communicative. However, most observers will know when such gesture is being performed. And this might be the real intention of the sender all along, for the observers to 'catch' his meaning. In such context, the action is intentionally communicative, and therefore is a gesture.

In HCI/computing literature, a pattern of agreement with Kendon's definition is that a behavior does have to be meaningful for it to be called a gesture, but instead of being meaningful to another person, it has to be meaningful to the computer. The definition that is widely circulated is proposed by Kurtenbach and Hulteen, which reads:

> *A gesture is a motion of the body that contains 'information'. Waving goodbye is a gesture. Pressing a key on the keyboard is not a gesture because the motion of a finger on its way to hitting a key is neither observed nor significant. All that matter is which key was pressed.* (Kurtenbach & Hulteen 1990, 310)

In comparing the act of waving goodbye with the act of pressing a keyboard, both are interactions, albeit one to another person and the other to a computer. But while the motion of waving can convey the meaning of parting, hence containing information, the motion of a finger reaching out to press a key does not hold any significant meaning to the computer. All that matters to the computer is which key is being pressed. Thus, in computing terms, a movement that *communicates* is the one that the computer process as input. The input is the user intent to make something happen (i.e., a command). And this intent is the "information" referred to in the definition above.

From this example, it is implied that how the act of pressing the key is done, whether with or without force, is not significant. The pressing of the key then could well be represented by such arbitrary acts such as rubbing one's eye, which, by Kendon's definition, is not a gesture. In the computing domain though, it might just be the 'gesture' that is recognized by the system to *mean* something.

Another point we can take from this definition is that the meaning of movements depends on those receiving it. Based on the idea described by Kurtenbach and Hulteen (1990) and Webb and Ashley (2012), this point can be described as the following. The above scenario of pressing a keyboard considers a one-to-one interaction, which means only one receiving end decides which movement is meaningful. Consider a scenario of two people (Person A and Person B) using one computer. In this interaction, we have a "speaker" (Person A) and two "listeners" (Person B and the computer). Person A is reaching out to press the key 'Y' on the keyboard of the computer while Person B is watching intently. The act of reaching out to press a key means that Person A is giving a command to the computer. Even though the act contains information (giving a command), it is only meaningful to Person B, and not (yet) to the computer. How the key is pressed also contains information, it could, for example, be pressed forcefully, but again this is only meaningful to Person B.

The moment a key is being pressed, it became meaningful to both "listeners": Person B *and* the computer. But, how the meaning is interpreted differs between the listeners. Person B interprets it simply as 'Person A is giving a command to the computer to type the letter "Y". This interpretation ends here without any further action needed for Person B. The computer interprets it as 'Person A is giving *me* a typing command by pressing the key "Y"' and consequently proceeds to output the letter "Y". Here, it can be said that the meaning is interpreted based on the *context* of that interaction.

Classifying Gesture

We turn our attention to the next fundamental aspect of gestures: their classification. Cadoz and Wanderley (2000), another prominent work in the HCI literature of gestural-based interactions, suggested analyzing and classifying gestures based on their functionalities, since a gesture is performed to fulfill a function or a purpose. Cadoz and Wanderley outlined three functional roles associated with gestures: epistemic function, ergotic function and semiotic function. We describe each of these functions based on our understanding of the literature reviewed, in the following.

Epistemic. Any movements that bring a bodily part into contact with the environment with the intent to sense and comprehend are of epistemic function (Cadoz & Wanderley 2000). These include tactile interaction or haptic exploration of the hands, feet and any part of the body. Familiar examples are stroking cloth materials to gauge its texture, or walking barefooted to appreciate the texture of grass or sand.

In addition to sensing the environment through touch, epistemic actions have also been designated to actions in which we alter our physical environment to gather information and facilitate cognition, as suggested by Kirsh and Maglio (1994). This type of action can be found in everyday activities such as separating objects by colors to gain information on how many of them are a certain color, or sorting out tools before starting a task, in order to reduce later search (facilitate cognition).

The most broadly defined concept of epistemic function is proposed by Magnani (2004), in which it includes "all actions that provide us with additional knowledge and information". An important part of this definition is that, unlike the previous two, it also includes actions that do not alter anything in the environment, and actions that do not need physical touch, such as "looking" from different viewpoints for the reason of checking or evaluating.

Ergotic. Movements that are used to change the state of the material/physical object are of ergotic function (Cadoz & Wanderley 2000). These are energy-transferring movements that apply forces, displacements and deformations to manipulate or create object/artifacts (Luciani 2007). An easy example is the movements of the hands sculpturing pottery out of clay. The fact that energy is exchanged is essential (Luciani 2007).

Semiotic. Movements that are used to convey meaningful information are of semiotic function (Cadoz & Wanderley 2000). Familiar examples are movements with communicative intent, such as the waving of the hand to convey farewell. Unlike ergotic functions, semiotic functions are gestures that aim fundamentally at transmitting information, and not energy, to the environment. That is the case of the gestures that accompany speech, of the sign language of the deaf-mute, and of the gestures of musical conductors, to name a few examples. Kendon (1988) distinguished these types of gestures along a continuum, known in the gesture field as Kendon's Continuum (McNeill 1992). Figure 1 shows the continuum as depicted by Donovan and Brereton (2005).

Figure 1. Kendon's Continuum

Based on the figure, the continuum represents the gestures' relationship with speech, with gesticulation being the most heavily depended on speech. McNeill (2006) further expands the concept of gesticulation to a classification scheme that consists of four more categories: iconic, metaphoric, deictic, and beat. Due to space limitation, we chose not to describe each of Kendon's Continuum of semiotic gestures and the categories of gesticulation gestures; however, the essence that we want to point to here is that from both of these perspectives, gestures are considered to be closely tied to speech and language.

The functional classifications of epistemic, ergotic and semiotic gestures could be said to represent the interaction of human with his environment. There are also existing classifications of gesture from the HCI perspective, i.e., for interactions between human and computer. Two highly cited classifications of Quek (2003) and Karam and Schraefel (2005) are especially prominent in the literatures of gestural-based interaction.

Quek (2003) proposed a "purpose taxonomy" which represents the human hands as a tool to accomplish three main purposes: to modify objects, to signal, and to complement the usage of language. Based on these three main purposes, gesture systems are divided into three systems: Manipulative, Semaphoric and Conversational. Karam and Schraefel (2005) extended these into several more subclasses. The following describes the main characteristics of each system.

Manipulative Gesture Systems. Systems that "permit direct manipulation of its entities", thus controlling the entity. Hand motion indicates the path or extent of the controlled motion; therefore, the gesture should provide parameters to the system that indicate the intent of the user's request to move, relocate or physically alter the digital entity (Quek 2003). One example is the work by Bolt (1980) in which a gesture – the pointing of the finger – are used to directly pick an entity and control its positioning, similar to the dragging of a file by a mouse cursor to a new location.

Semaphoric Gesture Systems. Quek (2003) defines semaphores as systems of signaling, where a number of 'whole' gestures are predefined as a set of signals to be communicated to the machine. On usage, users would be required to perform a signal (gestures) to be recognized by the system to determine which of the gestures in the predefined set is being performed. Each of the gestures can be attached with its own meaning, for example, the "victory" sign can communicate the meaning of "delete". The set of gestures may be either static gesture poses or predefined dynamic movements. Semaphoric Gesture Systems follow a certain sequence: Actor signals the gesture; the system processes the gesture and responds accordingly; the interaction either ends or restarts with another gesture pose (Quek 2003).

Conversational Gestures Systems. A system that processes conversational gestures does not require the user to perform any poses or to learn any gestures other than those that naturally accompany everyday speech (Quek 2003). Both speech and gesture interfaces are therefore essential in analyzing the hand movements in context of the user's speech topic to determine the meaning of the gestures (Quek 2003).

Finding a common ground

In this section, we discuss the connection between the two perspectives identified from the exploration of the definition and classification of gestures discussed above. We firstly highlight how both perspectives connect in the definition of gestures through the concepts of meaning and context, and secondly in the classification of gestures through the concepts of simultaneous and asynchronous modes of information transfer. These interrelations provide findings on 1) the importance of meaning and context in an interaction, 2) evaluating the information transfer mode in an interaction, and 3) the linkage from tasks to gesture function. Each of these findings is discussed in the subsections below.

Linking the definitions of gesture

Essentially, Kurtenbach and Hulteen (1990) equals interaction with the act of "giving a command". This contributes to the main differing point of how actions that are not, by Kendon's (1986) definition, expressive can still be called gestures. Granted, the two definitions were tailored for different interaction contexts; Kendon's is based on human to human interaction, while Kurtenbach and Hulteen's is based on human to computer interaction. Yet, we discovered two points of similarities, described as Finding 1 below:

Finding 1: The importance of meaning and context

- A meaning, whether linguistic-based or technical-based, must be attached to an action for it to be called a gesture.
- Regardless of how a gesture is linguistically or technically defined, it will still be interpreted based on the context that it is in.

Linking the classifications of gesture

The interaction between an "Actor" (the one producing the gestures) and his "Environment" involves emission and reception of information (Cadoz & Wanderley 2000). The course of this information exchange can alter both the informational and physical state of the Actor and his Environment. Epistemic function particularly represents several degrees of emission and reception of information. In the definition by Cadoz and Wanderley (2000) and Kirsh and Maglio (1994), the reception of information by the Actor from his Environment changed his informational state, that is, he is being informed (of how soft, how rough, et cetera) while simultaneously altering the physical state of that Environment. In contrast, for Magnani's (2004) examples of epistemic function: while the Actor still receives information from the environment, the physical state of the involving Environment does not change.

For ergotic function, it aims not only at informing the external world, but also at being informed by it (Cadoz & Wanderley 2000). More fundamentally, it also aims at transforming the world physically. Hence the simultaneous emission and reception of information changes both the informational and physical state of the Actor and the Environment.

Meanwhile, for semiotic movements, it changes the informational state of the environment (being informed) but not of the Actor. And unlike ergotic movements, the relayed information, while changing the informational state of the environment, does not change the physical state of it.

Finding 2: Evaluation of information transfer mode

The effect of each function to the physical and informational state of the Actor and Environment are summarized using Boolean logic in Table 1 below. During an interaction between an Actor and his Environment, if the informational state (A) of the Actor/Environment changes, it means that the Actor/Environment receives information. If the physical state (B) changes, it means that the Actor/ Environment emits information. Therefore, when evaluated using AND operator, the Q value denotes TRUE when there is a simultaneous emission and reception of information. Table 1 illustrates this concept.

State/	**Informational State (A)**			**Physical State (B)**		
Function	Actor	Environment	Q	Actor	Environment	Q
Epistemic	1	1 OR 0	1	1	1 OR 0	1
Ergotic	1	1	1	1	1	1
Semiotic	0	1	0	1	0	0

Table 1. Effect of Functions to Actor and Environment

Based on Table 1, we can then conclude that ergotic and epistemic movements allow both emission and reception of information to be done in a *simultaneous* transfer mode while semiotic movements only allow *asynchronous* transfer mode. We define this finding as Finding 2.

Finding 3: Linking task to function

Subsequently how does Finding 2 link with the HCI classifications of gestures? To answer this question, we will discuss the connection by traversing the evolution of interaction paradigms. We will start with the most prevalent: the WIMP paradigm. In a WIMP paradigm, when entities such as desktop folders are manipulated, the interaction allows simultaneous emission and reception of information – users emit information on how the object should be manipulated (e.g., click to select a folder and drag it to another folder), and receive simultaneous information on the object's movement (e.g., perceiving the object being dragged). There is a "feedback loop" mechanism in the interaction, allowing a simultaneous transfer mode of information. The hand movements used to control the mouse and keyboard, which are the most commonly used devices in a WIMP paradigm, also require energy transfer (pressing the keyboard, clicking the mouse). This shows that the movements of manipulative computing tasks are of ergotic function.

Manipulative computing tasks however are not only done through the use of mouse and keyboard. The emergence of post-WIMP interfaces allows manipulative tasks to be done through more advanced input devices such as the Head-Mounted Device (HMD), commonly used in virtual-based applications. The HMD directly manipulates the imaging system where the system reacts to the motion of the user wearing the HMD by producing a perspective change of the rendered image. A feedback loop mechanism similar to the use of mouse and keyboard exists, allowing a simultaneous transfer mode of information. The difference in this case of a HMD is that it only changes the perceivable state of the user (being informed) but not the physical state of the system environment. This shows that the movements (motions of the head) of this type of manipulative computing tasks are of epistemic function, as they agree with the previously discussed definition proposed by Magnani (2004).

The above described devices for manipulative tasks, i.e., mouse, keyboard and HMD, reflect the ongoing evolution of interaction paradigms. The evolution is now shifting the focus from the affordances of a device/interface to the natural affordances of the human body, i.e., body movements and gestures (without device), which is the central interaction theme of this paper. The following paragraphs discuss what kind of functional movements are required to control such gesture-based interfaces.

Reiterating the point made in Finding 2, movements of ergotic and epistemic function allow simultaneous emission and reception of information. This simultaneous exchange of information is what allows the tight feedback loop interaction essential in a manipulative task. Referring to Quek's classification above, this is exactly what Manipulative Gesture Systems represent, e.g., Bolt's (1980) "pointing" gesture to directly manipulate entities. Hence manipulative gestures, even without any use of devices, are still of ergotic and/or epistemic function.

Quek's classification shows that gesture interfaces include those that are not only used to manipulate objects, but also to "signal". As described earlier, these signals or semaphores are attached with meanings (information) to be "relayed" to the computer. In other words, it is a command and response system, which means the information is transferred in asynchronous mode. This concept alone allows for a direct relation to movements that are of semiotic function.

Quek's classification also shows that current gesture interfaces include those that are used to "converse" with the computer. Although the gestures are produced naturally instead of being predefined, they are still attached with "meanings" that require the system to process them before responding accordingly. In other words, Conversational Gesture Systems is still a command-and-response system similar to Semaphoric Gesture Systems in which information is transferred in asynchronous mode. Hence it is also directly related to semiotic function.

Paradigm	**WIMP**	**Post-WIMP**		**Gestural-based Interaction**		
Instrument	mouse, keyboard	glove	HMD	manipulative gestures	semaphoric gestures	conversational gestures
Main feature	energy transfer	feedback-loop interaction	feedback-loop interaction	feedback-loop interaction	signals	gesticulations
Goal	manipulate entities	manipulate entities	sense surrounding	manipulate entities	convey information	converse
Task type	manipulative	manipulative	manipulative	manipulative	communicative	communicative
Transfer mode	simultaneous	simultaneous	simultaneous	simultaneous	asynchronous	asynchronous
Function	ergotic	ergotic	epistemic	ergotic/ epistemic	semiotic	semiotic

Table 2. Mapping Task to Function

The link between psycholinguistic classifications and HCI classifications is thus depicted in Table 2 above, defined as Finding 3.

Towards Natural Interaction

From our observation, one of the main reasons for why the disparity issue exists is because the notion of how the psycholinguistic perspective of gestures i.e. classifications of Kendon (1986) and McNeill (2006) is said to be exclusive to speech-based gestures, and therefore deemed not suitable for HCI-based technologies that do not instill speech interfaces as an interaction modality. This drives most HCI researchers to decidedly adopt Kurtenbach and Hulteen's (1990) definition of gestures when developing gestural-based interfaces.

Regardless of the connections gesture has with speech though, the concept of *meaning/information* exists in both the psycholinguistic and the HCI perspective, as we have shown in our Finding 1. This shared perspective thus could justify that a gesture can indeed be "arbitrary/artificial" as long as a meaning/information is attached to it.

However, this does not mean that it warrants an excuse for choosing any random arbitrary/artificial gestures. If gestures were randomly chosen just for easy recognition purposes, it will render the system "unnecessary" as failed recognitions of gestures are problems that wouldn't be faced in traditional systems of mouse clicks and key(board) presses (Wexelblat 1998). It would take a justifiable reason to still choose predefined gestures for gestural-based system, and an even more justifiable mapping strategy to map the meanings to each gesture.

This is where our finding of the importance of context (Finding 1) could help. We reiterate that regardless of how a gesture is linguistically or technically defined, it will still be interpreted based on its context. Thus, when developing a gestural-based system/application, to avoid it being rendered "unnecessary", the gestures have to be derived and their meaning mapped based on *context*.

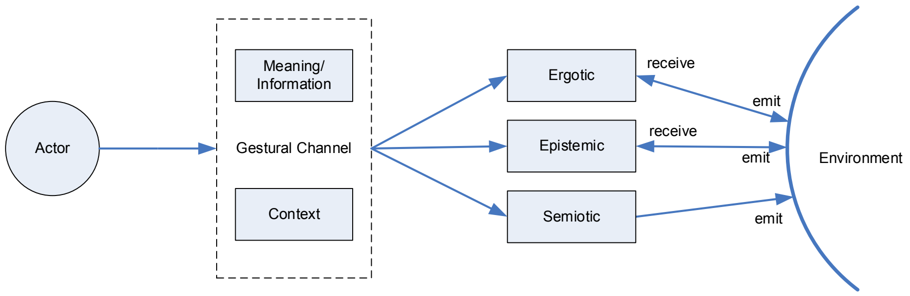

Figure 2. Conceptual framework of the findings

By analyzing the context of the problem that we want to solve using gesture-based systems, we should be able to identify the type of task needed, whether manipulative or communicative. Based on our Finding 3, we can then ensure that for each task that requires gestures, they are mapped accurately to the corresponding gesture functions. For example, gestures for a manipulative task should be derived by following the characteristics of epistemic or ergotic function, and gestures for a communicative task should be mapped following the characteristics of semiotic function. The findings can be summarized in a conceptual framework shown in Figure 2 above.

Conclusions & Future Work

The discussions in this paper have highlighted that psycholinguistics (human-to-human interaction and human–environment interaction) and HCI (human–computer interaction) shared an understanding of the importance of meaning and context when defining gestures. The two fields also connect in the classification of gestures through the functional classifications of ergotic (manipulative), epistemic (manipulative) and semiotic (communicative) movements. We demonstrated how these perspectives interrelate through discussions based on investigations of solid, established and influential works that form the foundation of gesture research in HCI. Thus, this paper sets the framework for future works that intend to propose gestures as the interaction modality for interactive systems.

It is important to note that this paper only focused on the definition and classification aspects, though we do believe that these aspects form the most fundamental aspect of gesture theory. In addition, these findings which serve as the starting point of our research, are purely conceptual, and therefore would benefit from further studies that can provide quantifiable indication of the significance of these connections.

Nevertheless, we believe that these findings could be especially useful to be applied with gesture-based systems built based on real world events or scenarios. Most involve natural body movements, such as simulation for triage/emergency training (Bartoli *et al.* 2012; Kato *et al.* 2016; Kayama, Kato & Okada 2015;), fire training (He *et al.* 2014), crisis management (Mora & Divitini 2014), religious training (Othman *et al.* 2015), and general collaborative tasks (Wang *et al.* 2017). These are typically systems where the advantages of using gestures seem apparent, but would require from developers the understanding of how humans communicate and interact with other humans in that event, and to a larger extent, the environment of the event. To translate such understanding into a gesture-based computing system, we would argue that by incorporating the knowledge of how the perspectives of human-to-human interaction, human–environment interaction and human–computer interaction interrelate, developers can ensure the gestures derived are the most usable to users.

Acknowledgments

This research was supported by the Fundamental Research Grant Scheme, Ministry of Higher Education, Malaysia (vot no.: R.J130000. 7828.4F865)

Nur Zuraifah Syazrah Othman is a senior lecturer and a member of ViCubeLab Research Group at Universiti Teknologi Malaysia (UTM). She has been working on the HCI field of natural user interface and is especially keen on exploring how gestures can be seamlessly deployed as a valuable addition to existing interaction techniques. She is a member of myHCI-UX, the Kuala Lumpur ACM SIGCHI chapter.
Contact: zuraifah@utm.my

Mohd Shafry Mohd Rahim is a professor and chair of the Office of Undergraduate Studies, Universiti Teknologi Malaysia (UTM). His research works covered mainly the domain of image processing, image data analytics, and computer graphics. Currently he is also involved in several human-centered research projects while being appointed as a research fellow at the Media and Game Innovation Centre of Excellence (MaGICX) and the Institute of Human Centred Engineering (iHumEn) at UTM.
Contact: shafry@utm.my

Masitah Ghazali is an associate professor and a member of ViCubeLab Research Group at the Universiti Teknologi Malaysia (UTM). Her interest in HCI had driven her to work in various domains; such as embedded systems, software testing, interactive technology, and with the recent project on behavioural change using persuasive technology. Recently, she has been appointed as an IT Academic Fellow at the Centre of ICT at UTM. She is one of the founding members and currently the chair of myHCI-UX, the Kuala Lumpur ACM SIGCHI chapter.
Contact: masitah@utm.my

Sule Anjomshoae received a B.Sc. degree in Graphic and Multimedia Software in 2012 and an M.Sc. degree in Computer Science from Universiti Teknologi Malaysia in 2014. She is currently pursuing Ph.D. at Umeå University, Sweden, working on generating and presenting intelligible explanations for actions and decisions made by artificial intelligence systems.
Contact: sule.anjomshoae@umu.se

References

ABNER, NATASHA, KENSY COOPERRIDER & SUSAN GOLDIN MEADOW (2015). "Gesture for Linguists: A Handy Primer." *Lang Linguist Compass* 9.11: 437-451.

ANGELINI, LEONARDO ET AL. (2014). "Gesturing on the Steering Wheel: A User-Elicited Taxonomy." *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA, September 17-19, 2014.* ACM, New York, NY, USA. 1-8

BARTOLI, GIOVANNA ET AL. (2012). "Emergency Medicine Training with Gesture Driven Interactive 3D Simulations." *Proceedings of the 2012 ACM Workshop on User experience in E-learning and Augmented Technologies in Education, Nara, Japan, November 2, 2012*. ACM, New York, NY, USA. 25-30.

BILLINGHURST, SABRINA & KIM-PHUONG L. VU (2015). "Touch Screen Gestures for Web Browsing Tasks." *Computers in Human Behavior* 53: 71-81.

BOLT, RICHARD (1980). "Put-that-there: Voice and Gesture at the Graphics Interface." *SIGGRAPH Comput. Graph.* 14.2: 262-270.

CADOZ, CLAUDE & MARCELO M. WANDERLEY (2000). "Gesture-Music." *Wanderley, M., Battier, M. (Eds) Trends in Gestural Control of Music.* IrcamCentre Pompidou. hal-01105543.

CHOI, SANGWON, ET AL. (2012). "Area Gestures for a Laptop Computer enabled by a Hover-Tracking Touchpad." *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction, Matsue-city, Shimane, Japan, August 28-31, 2012.* ACM, New York, NY, USA. 119-124.

DINGLER, TILMAN ET AL. (2018).“Designing Consistent Gestures across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses.” *CHI ’18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada, April 21 - 26, 2018*. ACM, New York, NY, USA. Paper 419.

DONOVAN, JARED & MARGOT BRERETON (2005). “Movements in Gesture Interfaces.” Critical Computing 2005 - Between Sense and Sensibility, the Fourth Decennial Aarhus Conference. Proceedings of the Worskshop : Approaches to Movement- Based Interaction, 21 August 2005, Denmark, Aarhus. IDWoP, UTS, Sydney. 6-10.

EVANS, ABIGAIL, JACOB WOBBROCK & KATIE DAVIES (2016). “Modeling Collaboration Patterns on an Interactive Tabletop in a Classroom Setting.” *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, California, USA, February 27 - March 02, 2016.* ACM, New York, NY, USA. 860-871.

GRANDHI, SUKESHINI A., GINA JOUE & IRENE MITTELBERG (2011). “Understanding Naturalness and Intuitiveness in Gesture Production: Insights for Touchless Gestural Interfaces.” *CHI ’11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, May 07 - 12, 2011.* ACM, New York, NY, USA. 821-824.

HE, QIUHAI ET AL. (2014) “An Immersive Fire Training System using Kinect.” *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, Seattle, Washington, September 13 – 17, 2014*. ACM, New York, NY, USA. 231-234.

HEIMONEN, TOMI ET AL. (2013). “Designing Gesture-Based Control for Factory Automation.” *Kotzé P., Marsden G., Lindgaard G., Wesson J., Winckler M. (eds) Human-Computer Interaction – INTERACT 2013. Lecture Notes in Computer Science, vol 8118.* Springer, Berlin, Heidelberg. 201-209.

KARAM, MARIA & M.C. SCHRAEFEL (2005). “A Taxonomy of Gestures in Human Computer Interactions.” <http://eprints.ecs.soton.ac.uk/11149> 1-45.

KASTENS, KIM A., LYNN S. LIBEN & SHRUTI AGRAWAL (2008). “Epistemic Actions in Science Education.” *Freksa C., Newcombe N.S., Gärdenfors P., Wölfl S. (eds) Spatial Cognition VI. Learning, Reasoning, and Talking about Space. Spatial*

Cognition 2008. Lecture Notes in Computer Science, vol 5248. Springer, Berlin, Heidelberg. 202-215.

KATO, RYUGA ET AL. (2016) "Proposal of an Architecture and Implementation of a Triage Support System." *Yoshino T., Chen GD., Zurita G., Yuizono T., Inoue T., Baloian N. (eds) Collaboration Technologies and Social Computing. CollabTech 2016. Communications in Computer and Information Science, vol 647.* Springer, Singapore. 246-261.

KENDON, ADAM (1986). "Current Issues in the Study of Gesture." *J-L. Nespoulous, P. Perron and A.R. Lecours, (eds) Biological Foundations of Gesture* Hillsdale, New Jersey: Lawrence Erlbaum, Associates, 1986. 23-47.

KENDON, ADAM (1988). "How gestures can become like words." *F. Poyatos (ed.) Crosscultural perspectives in nonverbal communication* Toronto: Hogrefe, 1988. 131-141.

KIRSH, DAVID & PAUL MAGLIO (1994). "On Distinguishing Epistemic from Pragmatic Action." *Cognitive Science* 18.4: 513-549.

KISTLER, FELIX & ELISABETH ANDRE (2013). "User-Defined Body Gestures for an Interactive Storytelling Scenario." *Kotzé P., Marsden G., Lindgaard G., Wesson J., Winckler M. (eds) Human-Computer Interaction – INTERACT 2013. INTERACT 2013. Lecture Notes in Computer Science, vol 8118.* Springer, Berlin, Heidelberg. 264-281.

KURTENBACH, GORD & ERIC E. HULTEEN (1990). "Gestures in Human Computer Communication." *Laurel, B., Ed., The Art of Human-Computer Interface Design.* Addison-Wesley Publishing Company, Inc., New York. 309-317.

LUCIANI, ANNIE (2007). "Ergotic/Epistemic/Semiotic Functions." *Enaction and enactive interfaces: a handbook of terms*. Enactive System Books. 96-97.

MAGNANI, LORENZO (2004). "Model-Based and Manipulative Abduction in Science." *Foundations of Science* 9.3: 219-247.

MCNEILL, DAVID (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

MCNEILL, DAVID (2006). "Gesture and Communication." *Encyclopedia of Language & Linguistics (Second Edition)*. Ed. Keith Brown. Elsevier: Oxford. 58-66.

MEDRANO, SAMUEL NAVAS, MAX PFEIFFER & CHRISTIAN KRAY (2017). "Enabling Remote Deictic Communication with Mobile Devices: An Elicitation Study." *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, Vienna, Austria, September 4-7, 2017.* ACM, New York, NY, USA. Article No. 19.

MORA, SIMONE & MONICA DIVITINI (2014). "Watchit: A Modular and Wearable Tool for Data Collection in Crisis Management and Training." *Aarts E. et al. (eds) Ambient Intelligence. AmI 2014. Lecture Notes in Computer Science, vol 8850.* Springer, Cham. 274-289.

NIELSEN, MICHAEL ET AL. (2003). "A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI." *Camurri A., Volpe G. (eds) Gesture-Based Communication in Human-Computer Interaction. GW 2003. Lecture Notes in Computer Science, vol 2915.* Springer, Berlin, Heidelberg. 409-420.

NORMAN, DONALD & JAKOB NIELSEN (2010). "Gestural Interfaces: A Step Backward in Usability." *Interactions* 17.5: 46-49.

NORMAN, DONALD (2010). "Natural User Interfaces are not Natural." *Interactions* 17.3: 6-10.

OTHMAN, NUR ZURAIFAH SYAZRAH ET AL. (2015). "Towards Movement-Based Interaction for Tawaf Training: Investigating its Suitability for Older Adults." *Jurnal Teknologi* 75.2.

QUEK, FRANCIS (2004). "Gesture Recognition." *Encyclopedia of Human-Computer Interaction, Vol. 1.* Ed. William Sims Bainbridge. Berkshire Publishing Group. 288-292.

RATEAU, HANAE, LAURENT GRISONI & BRUNO DE ARAUJO (2014). "Mimetic Interaction Spaces: Controlling Distant Displays in Pervasive Environments." *Proceedings of the 19th International Conference on Intelligent User Interfaces, Haifa, Israel, February 24 - 27, 2014.* ACM, New York, NY, USA. 89-94.

RUSNÁK, VÍT ET AL. (2018). "Designing Coherent Gesture Sets for Multi-Scale Navigation on Tabletops." *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada, April 21 - 26, 2018.* ACM, New York, NY, USA. Paper 142.

SHARMA, RAJEEV ET AL. (2003). "Speech-Gesture Driven Multimodal Interfaces for Crisis Management." *Proceedings of the IEEE* 91.9: 1327-1354.

SILPASUWANCHAI, CHAKLAM & XIANGSHI REN (2015). "Designing Concurrent Full-Body Gestures for Intense Gameplay." *International Journal of Human-Computer Studies* 80:1-13.

TAYAMA, YUKI, RYUGA KATO & KEN-ICHI OKADA (2015). "Triage Training System: Adjusting the Difficulty Level according to User Proficiency." *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia, Linz, Austria, November 30 - Dcember 02 2015*. ACM, New York, NY, USA. 139-147.

TSCHARN, ROBERT ET AL. (2017). "Stop Over There: Natural Gesture and Speech Interaction for Non-Critical Spontaneous Intervention in Autonomous Driving". *Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, November 13-17 2017*. ACM, New York, NY, USA. 91-100.

URAKAMI, JACQUELINE (2014). "Cross-Cultural Comparison of Hand Gestures of Japanese and Germans for Tabletop Systems." *Computers in Human Behavior* 40:180-189.

VAN DAM, ANDRIES (1997). "Post-WIMP User Interfaces." *Communications of the ACM* 40.2: 63-67.

VAN DAM, ANDRIES (2001). "User Interfaces: Disappearing, Dissolving and Evolving." *Communications of the ACM* 44.3: 50-52.

WANG, ISAAC ET AL. (2017). "Exploring the Use of Gesture in Collaborative Tasks." *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, USA, May 6-11, 2017*. ACM, New York, NY, USA. 2990-2997.

WEBB, JARRETT & JAMES ASHLEY (2012). *Beginning Kinect Programming with the Microsoft Kinect SDK*. Apress.

WEXELBLAT, ALAN (1998). "Research Challenges in Gesture: Open Issues and Unsolved Problems." *Wachsmuth I., Fröhlich M. (eds) Gesture and Sign Language in Human-Computer Interaction. GW 1997. Lecture Notes in Computer Science, vol 1371*. Springer, Berlin, Heidelberg. 1-11.

WOBBROCK, JACOB, MEREDITH RINGEL MORRIS & ANDREW DAVID WILSON (2009). "User-Defined Gestures for Surface Computing." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, April 4 – 9, 2009.* ACM, New York, NY, USA. 1083-1092.

WU, HUIYUE & JIANMIN WANG (2012). "User-Defined Body Gestures for TV-based Applications". *Fourth International Conference on Digital Home (ICDH), Guangzhou, China, 23-25 November, 2012.* IEEE. 415-420.