Journal of Visual Languages and Computing 25 (2014) 572–576

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

View Points

Gestures that people can understand and use ☆

Carmelo Ardito [a,*], Maria Francesca Costabile [a], Hans-Christian Jetter [b]

[a] Dipartimento di Informatica, Università di Bari Aldo Moro, via Orabona 4, 70125 Bari, Italy
[b] Intel ICRI Cities, University College London, Gower Street, London, UK

ARTICLE INFO

Article history:
Received 2 April 2014
Received in revised form
14 July 2014
Accepted 28 July 2014
Available online 21 August 2014

Keywords:
Gestural languages
Visual expressions
Symbolic gestures
Manipulations

ABSTRACT

Recent advances in computing devices push researchers to envision new interaction modalities that go beyond traditional mouse and keyboard input. Typical examples are large displays for which researchers hope to create more "natural" means of interaction by using human gestures and body movements as input. In this article, we reflect about this goal of designing gestures that people can easily understand and use and how designers of gestural interaction can capitalize on the experience of 30 years of research on visual languages to achieve it. Concretely, we argue that gestures can be regarded as "visual expressions to convey meaning" and thus are a visual language. Based on what we have learned from visual language research in the past, we then explain why the design of a generic gesture set or language that spans many applications and devices is likely to fail. We also discuss why we recommend using gestural *manipulations* that enable users to directly manipulate on-screen objects instead of issuing commands with *symbolic gestures* whose meaning varies among different users, contexts, and cultures.

1. Introduction

In recent years, the form factors of computers and the ways in which we interact with them have undergone a dramatic change. Three decades after the widespread adoption of graphical user interfaces with mouse and keyboard, there are now more smartphones in the world than there are desktop PCs. Interaction by touch is becoming the norm, rather than an exception. The ongoing advances in sensor and display technologies, CPUs, and wireless networks are a continuous source of innovation with novel devices ranging from very large displays to small wearables such as smart watches or augmented reality glasses. All these new devices will keep pushing researchers to envision new interaction possibilities that extend or completely replace traditional mouse and keyboard input for non-desktop scenarios.

One particularly successful example are large interactive displays or whiteboards that rely on pen or touch input and can now be found in many meetings and classrooms or even in our public spaces. Already very early attempts like *Liveboard* recognized the need for interaction beyond mouse and keyboard [1], since interacting with these devices felt comparably primitive and cumbersome and only very simple applications could be implemented. Over time, size and resolution of these displays increased and they also became far more affordable and widespread, so that the focus of large display development is now on complex applications that meet users' real-world needs in various situations. Therefore, in an attempt to make interaction more "natural", new modalities and interaction languages without mouse and keyboard are studied to improve the interaction with these new systems.

A particular area of interest is the use of human gestures and body movements. For instance, the presence of a human

☆ This paper has been recommended for acceptance by Shi Kho Chang

* Corresponding author.
E-mail addresses: carmelo.ardito@uniba.it (C. Ardito), maria.costabile@uniba.it (M.F. Costabile), h.jetter@ucl.ac.uk (H.-C. Jetter).

http://dx.doi.org/10.1016/j.jvlc.2014.07.002
1045-926X/

body in the proximity of a display can be detected by using different sensors (e.g. cameras, microphones, pressure sensors, Bluetooth, RFID scanners) to let the system automatically react to the presence or movement of users nearby. Ballendat et al. propose such a system as an interactive home media player on a large vertical display in a living room [2]. The system adapts displayed content and interaction possibilities based on proxemic information, i.e. distance, orientation, movements, and identity of people in relation to an ecology of multiple devices and objects in their nearby environment.

Today, thanks to advances in computer vision that permit real-time body, hand, and finger tracking, it is also possible to recognize human motion from a distance. Users can communicate with a system by performing a *gesture*, which in human–computer communication is defined as "a motion of the body that contains information" [3]. Gestures are claimed to enable a more natural and intuitive communication between people and devices. Ideally users do not think in terms of handling an input device, but naturally use their body to execute tasks or make use of their skills for gestural communication with other humans. The big challenge here is: how to design gestures that people can effectively understand and use?

Hand gestures, in particular, have been studied for a long time. One of the first papers on the topic was published in this journal in 1994 by Bordegoni [4]. It described a system supporting hand gestures for interacting with 3D user interfaces, which also provided a visual programming environment for the design of gestural languages that consisted of a set of hand gestures with each one containing information (as in the definition in [3]). Since then, however, only a few papers dealing with gesture design and use for human–computer communication have been published in JVLC. This is surprising, since, as we point out in the next section, gestural languages are indeed visual languages. With this article, we therefore hope to stimulate the research community interested in languages for gestural interaction and gestural user interfaces (not only limited to large displays) to consider this journal as an appropriate venue for their research.

2. Learning from visual languages

Let us recall the definition of visual language (VL) provided by the JVLC Editors in the foreword of the first issue of this journal in 1990: "By *visual languages* we mean the systematic use of visual expressions to convey meaning" [5]. The focus was on formal visual languages, which were studied with the goal of easing computer programming as well as human–computer communication through the use of graphics, drawings or icons. But this definition is also appropriate for natural and less formal languages. Examples of such more natural visual languages are the many different sign languages that are used worldwide to enable communication among deaf people or people who cannot speak: they use hand and body gestures to convey meaning. Similarly, gestures used by a human to communicate with a computer (performed either by the whole human body or by a part of it, e.g. the hands) are "visual expressions to convey meaning", i.e. visual languages.

We believe that for designing future visual–gestural languages for human–computer interaction we can learn from the past and capitalize on the great experience and lessons learned from three decades of VL work. 2014 marks 30 years from the first IEEE Workshop on VLs that was held in Hiroshima, Japan, in 1984. That workshop stimulated the research on VLs and started a series of workshops now held every year. In the mid-80s, the market availability of "high-resolution" graphical screens generated an enormous enthusiasm and the hope to greatly facilitate human–computer communication and programming by using VLs. The use of graphics promised to enable visual interaction by manipulating visual representations of objects and a better support of our human skills for visual information processing. In other words, it promised to solve an important real-world problem of the time by using newly available interaction technology, much like gestural languages for novel devices are now expected to solve interaction problems of our time. More concretely, the question of how to design a universal "standard set" of gestures is now a recurring theme in books, blogs, workshops, or special interest groups (e.g. see [6–9]).

Back in the 80s, one of the challenges of the VL community was to create visual programming languages that could be general-purpose, like Fortran or C. Their aim was to make programming easier for non-technical people. Glinert et al. addressed this challenge in [10], discussing several open problems that need to be solved to make this possible, for example creating a sound "graphical vocabulary", defining and validating metrics for assessing the relative merits of visual environments and programs, or developing scalable approaches. Other authors remarked the importance of finding new domains and various forms of visual languages where using graphics would be truly beneficial. Over the years, the idea of general-purpose visual programming languages demonstrated to be a failure. Visual representations have several advantages, but also many disadvantages: they are inherently ambiguous and often hard to understand or can only be interpreted within a certain context. Highly abstract concepts are too complex to be expressed visually. For example, many attempts have been performed to visualize *recursion*, but they resulted in very complicated images, difficult to understand. However, there are also success stories: the research on visual languages to facilitate database querying from the 90s (see [11]) resulted in visual interfaces which are much more usable than SQL for laypeople and are currently adopted by DBMS. Several domain-specific languages have been developed, which proved to be successful in practical applications [12].

3. General-purpose vs context-specific gestures

Can this experience inform on how we should approach designing gestural languages today? We believe it can. Like the VL community three decades ago, some researchers in HCI are now working on defining general-purpose gesture sets that are intended to be universally accepted by most people. To this aim, a specific meaning has to be assigned to each gesture. This results in *symbolic*