Touchless Gestural Interaction with Small Displays: A Case Study

Franca Garzotto
Politecnico di Milano
Via Ponzio 34/5 20133
Milano, Italy
franca.garzotto@polimi.it

Matteo Valoriani
Politecnico di Milano
Via Ponzio 34/5 20133
Milano, Italy
matteo.valoriani@polimi.it

ABSTRACT

Touchless gestural interaction enables users to interact with digital devices using body movements and gestures, and without the burden of a physical contact with technology (e.g., data gloves, body markers, or remote controllers). Most gesture-based touchless applications are designed for interaction with medium or, more often, large displays. Our research instead explores touchless gestural interaction "in-the- small", where the user interacts with small displays, of the size, for example, of a smart phone screen. Our work applies this to the domain of household appliances. We describe the design and evaluation of MOTIC (MOtion-based Touchless Interactive Cooking system), and highlight the complexity of employing touchless gestures to interact with visual interfaces constrained in size. This case study shows how existing design guidelines, mostly conceived for touchless gestural interaction "in the large", can be adapted and applied for gestural interaction in-the-small, and high-lights that touchless gestures have the potential of forming a valuable addition to our repertoire of interaction techniques with small displays.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: Multimedia Systems, User Interfaces

General Terms

Design, Experimentation, Human Factors.

Keywords

Touchless gesture interaction, motion-based interaction, design, small display, household appliance, Kinect, Design, Experimentation, Human Factors

INTRODUCTION

Touch-less gestural interaction enables users to explore multimedia information spaces or control digital devices using body movements without wearing additional aides (e.g., data gloves), being wired to a device (e.g., body markers), or handling remote controllers. This paradigm is expected to be more natural, spontaneous, intuitive, and pleasurable than other forms of interaction. The gestural dimension resembles one of the most primary forms of human-to-human communication – body expression. By their potential to engage the entire body, gestures can enhance the pleasure and engagement of participants, while the "come as you are" feature removes the burden of physical contact with technology, making the user experience more joyful.

CHItaly '13, September 16 - 20 2013, Trento, Italy

http://dx.doi.org/10.1145/2499149.2499154

Unlike touch gestures, touchless gestures have been implemented so far to a limited degree and in few contexts. In addition, most existing works assume that users interact with virtual objects, data, or digital functionalities presented in *medium* or *large* displays. Still, there are situations (e.g., in medicine, automotive contexts, manufacturing environments, or home) in which touchless gestures are more appropriate than the use of mouse or touch interfaces, e.g., because the user wears gloves, have dirty hands, or needs to operate at the distance or with free hands or in small spaces. At the same time, medium or large displays are not the right tool: they are too intrusive, do not meet the physical constraints of the space where interaction takes place, or do not fit into the objects involved in the user experience. In these circumstances, there is the need for touchless gestural interaction with visual interfaces fitting in *small screens* (of the dimension of smart phones). In other words, there is a need for *touchless gestural interaction "in the small"*.

Touchless gestural interaction in-the-small has received so far only a marginal attention in industry and research. Still, this paradigm has a potential in a number of interesting domains. In addition, it is challenging from a research perspective. Building any touchless gesture-based application has an intrinsic complexity related to both achieving accurate and meaningful gesture recognition and identifying natural, intuitive and meaningful gesture vocabularies appropriate for the tasks in question [18].

When dealing with touchless gestural interaction in the small, this complexity is exacerbated by the fact that gestures and their detection must be designed in relationship to visual interfaces that are constrained in size and have limitations on the number of interface elements and contents that can be displayed, on their dimension, and on the digital space available to provide feedbacks to user's interaction. This paper aims at providing a small step towards a better understanding of this complexity and how to master it. We present a case study in which we have applied touchless gestural interaction in the small to the domain of household appliances, and discuss the design and evaluation of a fully implemented touchless gestural interaction system called *MOTIC (MOtion-based Touchless Interactive Cooking system*). MOTIC comprises a conventional oven that is integrated with a small display and a Kinect motion sensing device (Fig. 1) to enable users control cooking behavior thought body motion and touchless gestures.

RELATED WORK

The analysis of existing applications of touchless gestural interaction pinpoint how this paradigm is mostly applied for interaction with large and medium size display, in a wide range of domains [4][29], mostly as proof-of concept research applications, while the use of this paradigm in combination of small displays is still very limited.

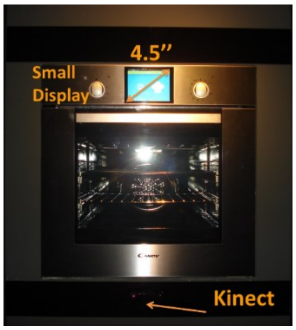


Figure 1. Front view of MOTIC with the integrated small display and Kinect sensing device

A number of gestural interaction systems are for entertainment (on platforms such as Sony Eye Toy or Microsoft Kinect). [19] for example describes a virtual game where game play is afforded by the user's silhouette interacting with on-screen 3D gaming objects. [30] presents a system - CoDine – that connects people in different locations through shared dining activities: gesture-based screen interaction, mutual food serving, ambient pictures on an animated tablecloth, and the transportation of edible messages. Rather than focusing on functionality or efficiency, CoDine aims to provide people with an engaging interactive dining experience through enriched multi-sensory communication. Other applications can be found in public shared interaction spaces [4]. [15] discusses touchless gestural interaction with framed digital displays integrated with walls and facades that creatively motivate individual and group interaction. The authors introduce the concept of "stage" as an apt metaphor to explore the ways in which these ubiquitous screens can transform passive viewing into an involved performance. [6] exploits cylindrical displays as a possible form of novel public displays which are especially suitable to keep people in motion and to support gesture-like interaction. In consumer electronics, touchless gestural interaction can to resolve the complications of using different remote controls. [26][28] present a review on gesture control for consumer electronics devices, introduces different sensing technologies and then focuses on camera-based gesture sensing and interpretation. [27] exploits touchless gestural interaction in the context of shopping activities. During sales conversations in a conventional shop gestures and mimics are of high importance to communicate information about a product. This paper analyzes one scenario in which an interface based on touchless gestures can support this form of communication between customer and employee. Medicine is a particularly promising domain for touchless gestural interaction [1][13]. Computerized medical systems play a vital role especially in the operating room; still, sterility requirements and interventional workflow often make interaction with these devices challenging for surgeons. Typical solutions, such as delegating physical control of keyboard and mouse to assistants, add an undesirable level of indirection. Touchless gestural interaction allows medical operators in operating rooms accessing information while maintaining total sterility. Other works explore contexts where motion-based touchless interaction with large displays facilitate collaborative exploration of large volumes of complex information (e.g., in control centers for crisis management and disaster relief, stock exchanges, manufacturing control plants). [17], for example, uses touchless gestural interaction for controlling 3-D virtual globes such as Google Earth (including its Street View mode), Bing Maps 3D, and NASA World Wind.

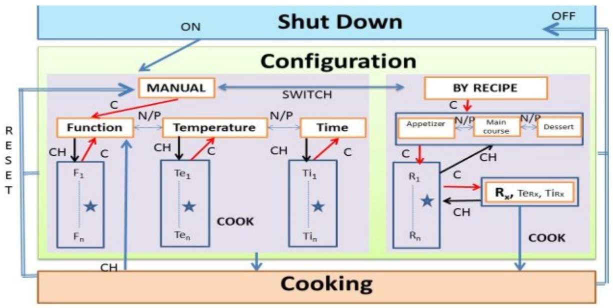
All the above mentioned systems exploit touchless gestural interaction assume that users interact with medium or large displays. The few existing applications that involve small displays are found in robotics - where remote gestures are used to give remote commands to small mobile robots (e.g., [5]) and above all in the domain of automotive, where touchless gestural interfaces are used to control and monitor vehicular systems while driving [23][24][9]. The primary motivation of research into the use of remote hand gestures for in-vehicle controls is broadly based on the premise that the risks in safety induced by taking the eyes off the road to operate conventional car controls can be reduced by using hand gestures. [9] investigates if touchless gestural interaction can facilitate drivers' control of car infotainment systems as well as comfort functions. It presents two user studies conducted with a working prototype of a multi-touch steering wheel, which pinpoint that driver's visual demand is reduced significantly by using gestural interaction. [23] provides a summary of current automotive hand gesture recognition research, and then compares touchless hand gestures versus conventional user controls, concluding that the former may offer, potentially, higher safety benefits without compromising the function of driving.

Most of the above cited works pinpoint the complexity of the design process in touchless gestural applications,, mostly in relationship to the definition of the gesture language. From a methodological perspective, touchless gestural interaction has attracted the attention of researchers since two decades [16]. [3], which dates back to the nineties, provides one of the first structured sets of design principles for free-hand interfaces. [7] offers some theoretical underpinnings for classification of gestures, in terms of their communication function, and highlights the role of culture and context in gestural languages. Other classifications of human gestures are discussed in [8] [25]. Many authors have identified heuristic and guidelines for touchless gestures design, either by elaborating the experience gained by developing prototypes, or by performing user experiments and behavioral observations to identify which are the most natural and intuitive human gestures for some categories of human tasks [8][14][20][31]. Attempts in this direction have been also undertaken in industry contexts (e.g. [11]). Other works have proposed tools to facilitate design, implementation, and evaluation of gestures [2], or have suggested procedures to find and test the specific gestures that make interaction more efficient, usable, and ergonomic [20]. Differently from (multi)touch interaction, in which some standardized gesture sets exist, there is no widely accepted set of touchless gestures nor a shared methodology for creating a touchless gesture set. The debate about the naturalness of existing gestural languages is still open [10][18][20][21][22].

MOTIC REQUIREMENTS

MOTIC has been developed in the context of a project sponsored by Candy–Hoover SpA, the national branch of a large multinational industrial group that plays a leading role in the domestic appliance sector worldwide. The project implements the company's strategy of "meeting the competitive challenge through innovative products" (http://www.candy-group.com), and has two main goals: i) exploratory research: investigating the exploitation of novel interactive technology for household appliances, identifying its effectiveness for customers, and understanding its industrial feasibility, to orient future decisions at a more strategic level; ii) branding & marketing: MOTIC has been built for being demonstrated in Mid April 2012 at the largest worldwide fair in the Home Furnishing Sector and installed in Candy-Hoover show rooms.

The requirement elicitation activity has involved our research team and a group from Candy-Hoover composed of the marketing general manager, the cooking division product manager, an industrial designer, a product developer, and an ICT engineer. Various types of requirements have been identified. Integration requirements have defined the key constraints on the physical installation of the system components. MOTIC must be consistent with the conventional way of integrating "normal" ovens within home kitchen furniture. Hence: i) the small screen must be an integrated component of the oven itself and have the same size (4.5"-5") as the displays which are currently mounted on top-level products to show cooking functions and parameters; ii) the oven must be fitted into a kitchen unit - an upright cupboard; iii) the Kinect must be hidden inside the furniture. Considering the most standard oven positions in a cupboard, the sensing device must be placed at the approximate height of 1 meter at the bottom of the oven, at the approximate height of 1 meter. In terms of functional requirements (Fig. 2), MOTIC must support all commands and controls that are available on Candy top-level ovens by means of buttons and knobs. These ovens offer two cooking modes - manual and "by-recipe".



Caption: F= oven Function; Te= Temperature; Ti= Time; R= Recipe; C= select& Confirm; CH= Change; N/P=Next/Previous; ★=explore values

Figure 2. MOTIC Functional Requirements

In the manual mode, the user sets the cooking parameters – "function" (e.g., "natural convection", "grill", "fan-cooking"), time, and temperature. In the "by-recipe" mode, the user selects a type of recipe (e.g., "main course", "meat", "fish", "vegetables", "dessert") and then a specific recipe (e.g., the "main course/lasagna") among the ones on display. In both modes, the user must press a start button to launch cooking, which takes places according to the user-defined parameters or the built-in parameters for the chosen recipe. Finally, in terms of lay-out requirements, the visual interface must be coherent with the "brand image", i.e., it must employ symbols, colors, and shapes as in conventional Candy-Hoover oven displays and product manuals.

THE DESIGN OF MOTIC

The design of any system based on gestural interaction involves two main dimensions. One concerns the definition of the gesture language, which specifies the shape (trajectory) and dynamics of body movements that express what the user wants to do and communicate to a system while performing the tasks identified during requirement elicitation. The second design dimension is related to the specification of the visual interface on display, presenting the information and functions which the user can operate on and the feedbacks on interaction. Finally, in applications requiring a smooth integration of the gestural system with the physical environment, a third design dimension needs to be addressed, concerning the actual positions of the hardware devices (sensing device and display). This dimension, in the case of MOTIC, is related the actual position of the Kinect within the home kitchen furniture (being the display position fixed).

To address the three design dimensions, we have followed an iterative user-centered design process during which we have repeatedly defined, implemented, and tested gestures, visual interfaces, Kinect positions using progressively more sophisticated artifacts (from paper based mockups to functional prototypes) and involving users along the entire process.

Initial Design: Paper Based Protyping

We built a paper based prototype of the visual interface and involved 16 users (6 from Candy-Hoover and 10 from our department, aged between 26 and 49) to identify a preliminary set of gestures. The paper mockup was composed of a set of posters, each one presenting a scenario of interaction with the application. A scenario corresponds, conceptually, to a sequence of application states (i.e., a path in the oven functional model of Fig. 2) and was rendered to the user as a sequence of interface sketches, each one visualizing an application state and having the same dimension as the application display. Users had never experienced touchless gestural interaction. They were asked to simulate the interaction with the application, performing one or more gestures that they felt appropriate to trigger state/sketch transition, thinking aloud during the execution of these tasks, and giving explanations of why they performed the way they did. In spite of some degree of variability among the users' gestures, we could identify some gestural patterns. Based on our observations and the analysis of video recordings of each session, we derived our initial gesture language, which was progressively refined in the following design activities.

Intermediate Design Phase: Functional Prototyping

Based on the results of the initial design phase, we defined a preliminary gestural language and built a set of progressively more complete functional prototypes, where the small screen, the Kinect, and the computer where mounted on the back of a cardboard structure showing a printed realistic 2D rendering of oven (Fig. 3). We performed a formative involving a larger group of users (30 persons recruited at our department - faculty members, students, and administrative staff. The functional prototypes were integrated with a tool for gesture monitoring (Fig. 4).

Figure 3. MOTIC Functional prototype

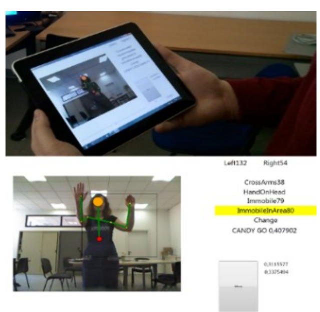


Figure 4. The gesture monitoring tool

The tool was created to meet the needs of iterative gesture design as outlined in [2]. First, we wanted to support iteration by making it easy to explore alternatives. The tool allows for quickly try out different motions for a gesture-activated function and for experiment with different recognition parameters in order to get the desired results. Second, we wanted to support retrospection, to help designers recall and examine the gestures they've created, understand how they are similar or different from other gestures. Finally, we wanted to support further testing and redesign, if needed after summative evaluation.

The monitoring tool provides, on a tablet PC (Fig. 4), a dynamic visualization of the user and her body skeleton, of the gesture that is detected and recognized by the system and how it is interpreted in terms of gesture detection and effects on the visual interface; it is integrated with the code editor to allows for a contextual, on-the-fly modification of detection parameters if needed. All recorded data are stored in a structured log file. Hence the tool enables the evaluator or the developer to monitor MOTIC behavior in real time, to check the precision and accuracy of the system as user interaction proceed, to immediately detect and correct potential errors, and to perform for further off-line analysis.

Final Design

This section presents the MOTIC gesture language that was implemented in the final version of the system. We analyse our solutions in light of existing guidelines ([3][15][21]) and discuss the motivations of the design choices taken to meet the requirements of our system and the constraints of touchless gestural interaction *in-the-small*.

MOTIC final gesture language comprises 6 gestures, involving the upper part of the body only to meet the requirements discussed in the previous section (see Fig.s 5 and 6). The meaning of a gesture (i.e., its effect on the application state) depends on the current execution context. For example, "raising right hand" (Vertical Browsing) determine an increase of time or a scrolling of the menu of recipes (Fig. 6.6), depending on the application state and what is displayed on the screen.

Guideline 1: Semantic intuitiveness. Gestures should have a clear cognitive association with the semantic functions they perform and the effects they achieve.

MOTIC solution. User's tasks that trigger state transitions, can be classified in two categories: issuing a state-control-command (start/end cooking, shut-down, restart, set-configuration) and exploring a set of "values"– oven functions, temperatures, or time durations, or recipes. We used this distinction to classify gestures, make them more intuitive, and help users to grasp meaning of gestures more easily. Different types of movements are defined for the two classes. "Continuous" and fluid gestures suggest the idea of "browsing" or "surfing" and are particularly appropriate for exploration of large sets of items and for interacting with large set of values having a continuous nature (like time). Continuous gestures are Vertical browsing - raising arm to indicate "increase temperature (or time)" or "show more recipes", and Horizontal Browsing, to switch among horizontal visual elements). "Control" gestures are more rigid, net and fragmented. They are associated to interaction with discrete sets of values, and, because they suggest the idea of "command", they are associated to control functions such as Confirm or Back and Undo.

Guideline 2: Minimize Fatigue. Gestural communication involves more muscles than keyboard interaction or speech.

Gestural commands must therefore be concise and quick, and minimize user's effort and physical stress. Two types of muscular stress are known: static, the effort required maintaining a posture for a fixed amount of time; dynamic, related to the effort required to move a portion of the body through a trajectory. Hence the design of gestural interactions must avoid gestures that i) require high precision over a long period of time and cause an intense muscle tension over long periods (a syndrome commonly called "Gorilla arm); ii) span across a wide space and involve long trajectories.

MOTIC solution. MOTIC gestures have been defined taking into account that each body joint is characterized by a maximum mobility range, but the "comfort" range for the user is much smaller (Fig. 6). A movement trajectory of MOTIC gestures always falls within the joint comfort range. The *widest trajectories* – traversing the entire comfort range - are associated to explorations of *large amounts of values*. For example, Vertical Browsing, i.e., continuous vertical arm movements is used to increase and decrease time and temperature not only because they are more intuitive (see previous guideline) but also because the vertical mobility of arm joints is wider than horizontal mobility. For the same reason, Vertical Browsing is also used to scroll the set of receipts (which include over 30 items). Horizontal Browsing, i.e., horizontal arm scrolling is used to explore *small sets of values* (e.g., the set of oven functions). In addition, exploration of a large set of values is associated to "continuous" gestures, to avoid the effort of many repeated movements. Fragmented "control" gestures, more tiring, are used for interacting with small sets of values.

*Guideline 3: Learnability***.** It must be easy for the user to learn how to perform and remember gestures, minimizing the mental load of recalling movement trajectories and associated actions. The gestures that are most natural, easy to learn and are immediately assimilated by the user are those that belong to everyday life, or involve the least physical effort. These gestures should be associated to the most frequent interactions.

MOTIC solutions: The most frequent task is "make a selection" within a set a parameters or values. The associated action is the conventional "stop" gesture ("Confirm"). Other frequent gestures are those to increase/decrease temperature and time, which correspond to Vertical Browsing.

Guideline 6: Feedbacks. Appropriate feedback indicating the effects and correctness of the gesture performed is necessary for successful interaction, to improve a user's confidence in the system, to allow users to learn the appropriate manner of performance and to understand what was wrong with their action.

MOTIC solution. While feedback visualizations must maximize intuitiveness and expressiveness, it is very difficult give good visual feedbacks that are both adequate to the wideness (or height) of gestures and fit in a small screen. For layout constraints and precision reasons (see discussion in Guideline 5) each screen shot can display only very few elements. Adding elements that represent dynamic feedbacks to the actual contents of the application may impose a shrinking of *all* visual elements, which can make the screen unreadable and so reduce precision.

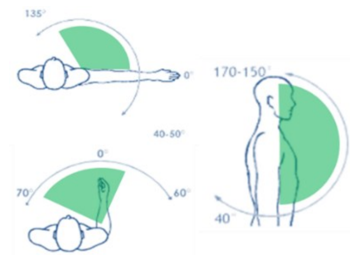
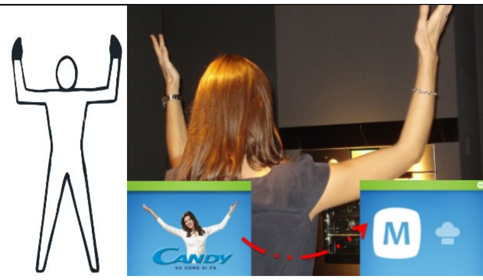


Figure 5. Gesture design: comfortable movement range of joints (in green)

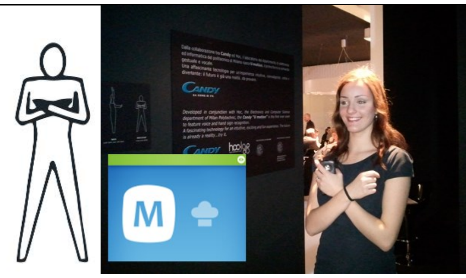
In MOTIC, feedbacks are provided in the following circumstances: i) to notify that a gesture is being traced by the system (the "eye" icon appears – Fig. 7); ii) to provide visual cue for the selected item among those on display (the corresponding visual element is highlighted); iii) to indicate the need for staying still for a while in order to "confirm a selection", i.e., to have the system interpret the stand behavior as intentional (progress bar–Fig. 7).

Guideline 7: Provide reversible actions. Commands must be easy to undo, to easily cancel any unintended action; "backward navigation" must be supported, to allow user return to previously seen objects or revise previous choices. *MOTIC solution.* Two gestures have been defined to enable for reversible actions: *Back/Undo* - moving left hand over the head (Fig. 6.5) to cancels the effects of last action-chose as it resemble "throwing something away", and *Reset* - crossing arms (Fig. 6.4), a shortcut to cancel all effects of interaction (parameter settings and recipe choice) and return to the start state.The latter gesture has been chosen as it evokes the X symbol of "close/undo" in many GUI interfaces

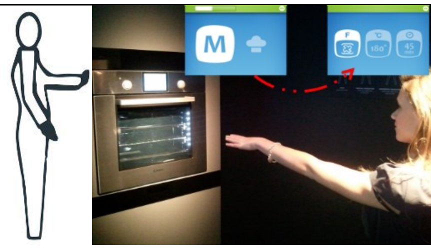
Guideline 4: Intentionality (Immersion Syndrome). Users can perform unintended gestures, i.e., movements that are not meant to communicate with the system they are interacting with. These are usually evoked when the user is communicating simultaneously with other devices or people, or just resting his or her body. The "immersion syndrome" occurs if *every* movement is interpreted by the system, whether or not it was intended, which may determine interaction effects against the user's will. The designer must identify well-defined means to detect the intention of the gestures, as distinguishing useful movements from unintentional ones is not easy. Body tension and non-relaxed posture of users (e.g. vertical full extension of arms joints) can be used to make explicit the user intention to start interaction, issue a command, or confirm a choice (in the same way that intonation distinguishes ordinary conversation from imperative orders). The tense period should be short to not generate fatigue.



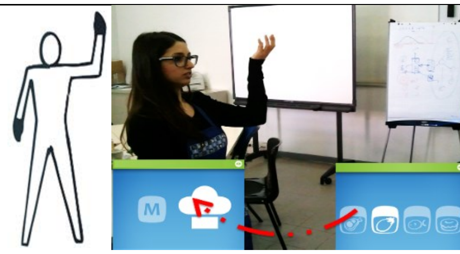
6.1: START/END COOKING: Turning oven "ON" or "OFF"(notice Candy-Hoover logo on the right screen)



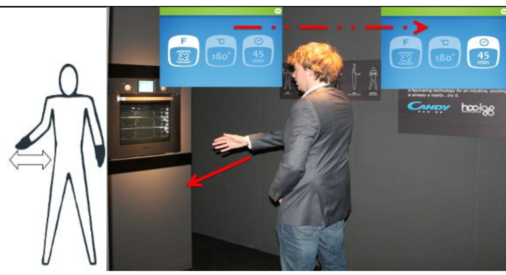
6.4: RESTART (e.g., to reset the cooking parameters): crossed arms



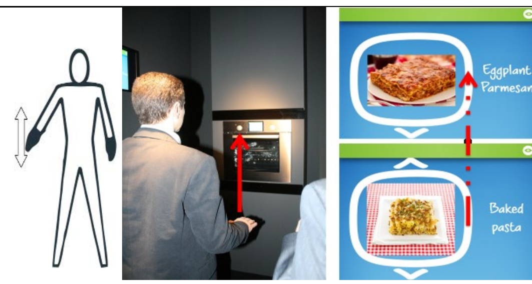
6.2: CONFIRM a selection (e.g., a cooking mode)– left arm steady



6.5: UNDO (e.g., to go back to a previous exploration step) – left arm towards the back



6.3 HORIZONTAL BROWSING: moving arms left/right to pick up a function



6.6 VERTICAL BROWSING: moving arms up/down to look at recipes-not intuitive by undrestood

Figure. 6. MOTIC Gestures and their corresponding visual interface.
Two scenarios (manual mode – left, and by recipe mode - right)

MOTIC Solution. To deal with intentionality, we have combined two concepts: "still" behaviors and "active zones". We assume that when a person suspends a movement and stays immobile for some seconds, it is likely that she is underlining her current behavior and intention. However, it may happen that a person gets to stand still unintentionally. To reduce the risk of misinterpretation, the second factor that contributes to establish intentionality is the space where the still gesture takes place ("active zone"). For example, the intentional selection of a value in a large set is associated to the Confirm gesture " keep left arm still *for 1 sec"* within the active zone "parallelepiped 30 cm distant from body, centered on belly, wide twice the shoulder width".

Guideline 5: Precision. Tasks that require precise interaction, e.g., fine selection of a specific value in a large set of alternatives presented on the screen, may be problematic: when operating at a distance, we cannot obtain a good resolution because of the intrinsic instability of movements in the free space. Touchless gestural input or control should be carefully designed with a special attention to precision.

MOTIC solution. Settings of oven parameters is an example of "task that requires precise interaction": users must make selections within fine grained sets of values - temperature 30°- 250° and time 5'-180' - that are quite large even with discretization. Using large displays, a similar problem can be alleviated in several ways, e.g., splitting a large set into multiple subgroups, and mapping each item on a large shape on the screen. When working with small displays, precision problem cannot be addresses just by group splitting and plain stretching of the size of graphical objects. Our tests have shown that for achieving an acceptable level of precision, only few values (maximum 6) should be displayed on a 4"-5" screen. For large sets of values, the number of sub-groups would be huge, introducing the extra burden of navigating across them. The MOTIC solution was: i) using continuous vertical gestures and maximizing the width of the trajectory (within the comfort range – see Guideline 2); ii) segmenting the continuous stream of captured motion into discrete lexical entities, and discretize the value set accordingly; iii) for each discrete movement detected, displaying one corresponding value on the screen.

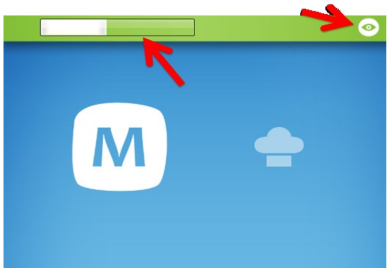


Figure. 7. Feedbacks cues in MOTIC

Guideline 8: Reconfigurability. A touchless gestural system is intended to be used, in principle, by many different types of users, who may execute the "same" movement in different ways, e.g., because of their different habits or anthropometric characteristics. The design trade-offs are between reducing technical complexity - by instructing the user and forcing him or her to execute a given gesture in a very specific way - and increasing technical complexity to make the system adaptable by adding customization functionalities. Within the second choice, customization should avoid overwhelming the users by requiring exhausting "calibration" activities (for automatic customization) or selections from long lists of tunable parameters and menu.

MOTIC solution. In MOTIC there is no need for user's calibration activity or setting of gesture detection parameters. The system exploits user's head, hands, and center of mass data *relative to* user's height in order to recognize gestures (see details in the next section). Using joints relative values increases the implementation complexity but supports automatic adaptation to morphologically different subjects without extra burden on the user. The only constraint for the user is that she or he must stand at a minimal distance from the sensing device. When the user is too close to MOTIC, a sound alert is generated.

IMPLEMENTATION

MOTIC implementation exploits the latest version of Kinect technology [11], including the Kinect for Windows sensing device and the Kinect Microsoft SDK API version 1.0 software libraries [12]. When the user enters the sensor range, the built-in skeleton tracking component of Kinect SDK processes the depth image data to establish the positions of various skeleton joints on a human form, and builds a skeleton representation. For example, the skeleton tracking determines where a user's head, hands, and center of mass are, and provides the 3D coordinate for each of these skeleton points. Such information is used by the MOTIC system for identifying and interpreting the user' gestures by means of two main components: the *GestureRecognizer* and the *GestureManager*.

The GestureRecognizer dynamically analyses a subset of all body skeleton data detected by the sensing device (considering only information related to torso and upwards joints) and progressively checks if such data correspond to a "legal gesture", i.e., meet the specifications of one of the gestures defined in MOTIC gesture language. Each gesture is modeled by a 2D active zone, a start condition, a set of constraints, and an end condition. For example, a user "horizontal wipe of the right hand" movement is interpreted as "Horizontal Browsing" if i) the right hand is detected in the area identified by the horizontal line "from left shoulder to twice the distance among shoulders" and the vertical line from "from left shoulder to the center of mass (belly) line" (active zone); ii) "the right hand is moving horizontally" (start condition); iii) "the motion continues with minimum speed V" (constraint); i) "the traversed distance is ≥D" (end condition). The GestureRecognizer is fully parametric so that the designers can easily change the setting of gesture specification parameters to adapt them to the anthropometric characteristics of a specific user. Once a gesture has been recognized, its data are passed to the GestureManager for interpretation and execution. As mentioned in the previous section, the semantics of a gesture (i.e., its effect on the application state) depends on the current execution context. The GestureManager checks for gesture intentionality, i.e., if the current gesture is meaningful in the current context, and only in this case it executes the corresponding application state transition (otherwise the gesture is ignored), generating the proper content on the small display.

FINAL EVALUATION

Research variables

The goal of the summative evaluation was to validate our design choices and to assess the design quality of MOTIC in terms of the following factors: i) the *quality of the gesture language*, operationalized in terms of 4 attributes related to gesture execution *(intuitive, easy-to-learn, simple, tiring*) and 2 attributes – *appropriate behavior* and *appropriate visual feedback* in response to a gesture – respectively related to the semantics and precision of a gesture, and to the correctness of its feedback; ii) the *global quality of the user experience* with MOTIC, operationalized in terms of an 5 attributes (*fun, intuitive, natural, irritating, incomprehensible, tiring*). In addition, we explored the *utility* of touchless gestural interaction, defined as the degree at which users would like to have this paradigm available in contexts different from gaming and specifically on their oven.

Procedure

We used a questionnaire to assess the above factors. For each design quality attribute, users were asked to express their degree of agreement/disagreement, on a 5 level Likert scale (1= strongly disagree; 5= strongly agree) rating the degree of agreement/disagreement that an attribute is to be associated to a specific gesture or to the overall experience.

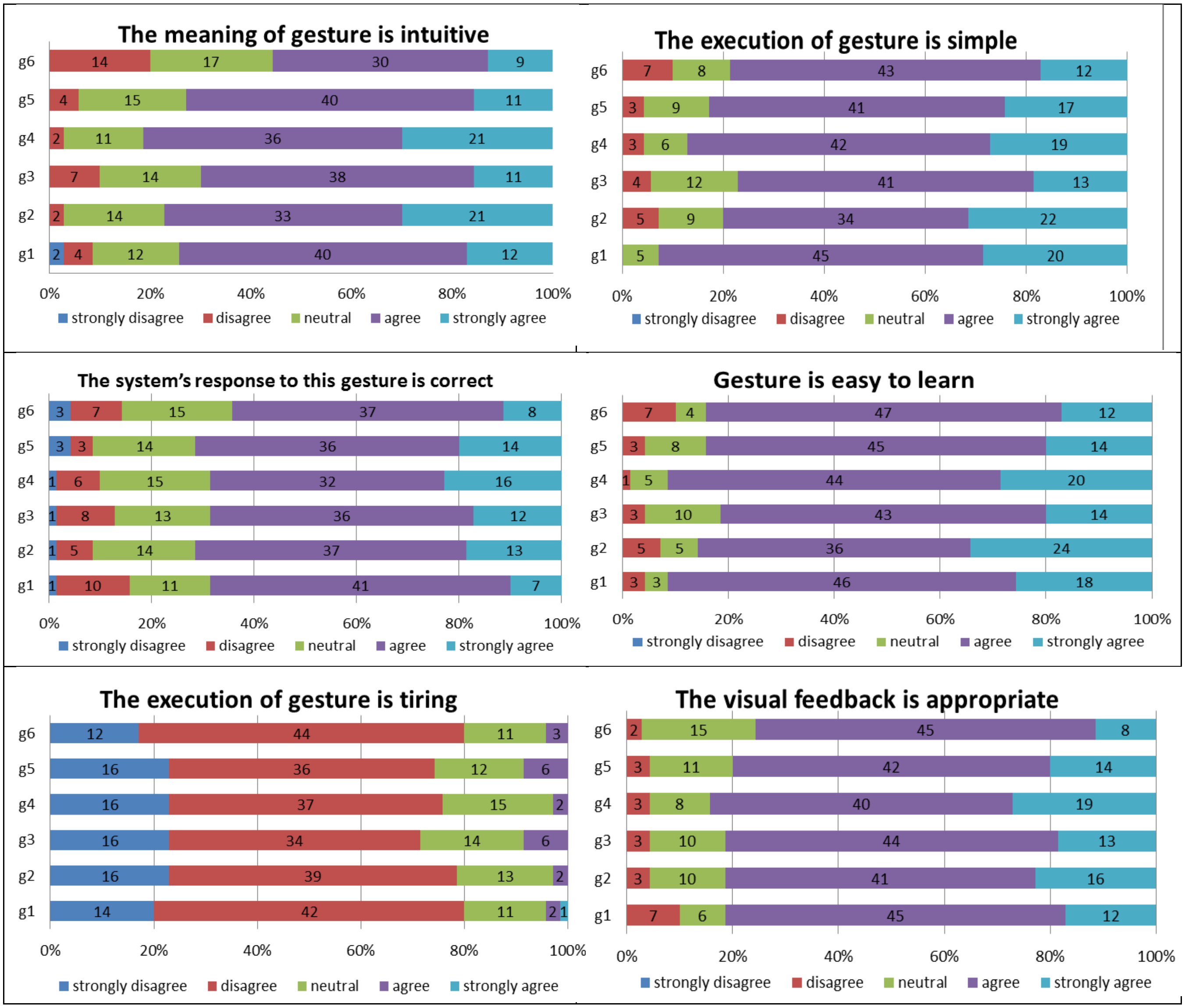


Figure 8. Overview of summative evaluation results

The summative evaluation activity was performed in mid April 2012, for one week at the Candy-Hoover stand at "Salone del Mobile" (involved 1.229 exhibitors from all over the world, and was attended by 286.391 visitors - http://www.cosmit.it/), the largest and more important commercial fair the Home Furnishing Sector and related sectors (e.g., home appliances). *70* participants were recruited among the people visiting the company stand. In each testing sessions, we fist provided a short explanation of the system and the gesture language for approximately 3 minutes (using a poster displaying the gesture language). Then we invited the user to perform the following tasks, defined in order to have the user experiment all gestures: "cook in manual mode" (i.e., setting the three cooking parameters); "reset the system"; "cook in by-recipe mode" (i.e., exploring available recipes and selecting a specific one). The average duration of a session of use was 5-7 minutes (as most users, after performing the tasks, kept playing with MOTIC for a longer time, often in group).

The respondents were 62,9% male and 37,1% female, 37,1% aged 20-30, 32,9% aged 30-40, 25,7% aged 40-50, and 4,3% aged 50-60. 24,3% of respondents had never experienced motion based touchless interaction, 64,3% had seldom used it (2-5 times), while 11,4% uses it frequently (2-10 times per month), all of them for entertainment purposes (employing technologies such as Wii and Kinect).

Results

The following diagrams shows the main results of summative evaluation, presenting the aggregated data related to user's rating of the different attributes for each gesture: Start/Cook (g1), Horizontal Browsing (g2), Confirm (g3), Vertical Browsing (g4), Reset (g5), Back/ Undo (g6).

Concerning intuitiveness, all gestures were perceived as intuitive or very intuitive by more than 80% of participants, and the number of negative responses is relatively low. The most intuitive gesture is “Vertical Browsing” (g4 – arm up/down continuous movement) followed by “Horizontal Browsing” (g2 - arm left/right continuous movement). The perceived quality of these gestures is also confirmed by their values in relationship to fatigue, precision, and semantics, highlighting that overall they can be considered as the most natural gestures. Indeed, both of them belongs to the human gesture vocabulary with a similar meaning in most cultures. The third most intuitive gesture is “Confirm” (g3 – arm still). The *less intuitive gesture* is “Start/Cook” (g1), which, admittedly, was mainly selected from marketing purposes: it mimics the company logo (see Fig. 6.1) that appears in all products and branding actions. Hence, it is semantically related only to the Candy-Hoover brand image and has no direct maching with its function. Still, “Start” is the most effective gesture in terms of learnability, followed by “Vertical Browsing” (g4). “Confirm” (g3) is perceived as intuitive - probably because it corresponds to the standard confirmation gesture of most videogames and to the conventional “stop” gesture, but it results the most difficult to learn, in spite of its intuitivennes. The weakness of this gesture is also confirmed by the low values of simplicity, and the high values of fatigue and correctness. User observations during the evaluation sessions pinpoint that many users attempted to express a confirmation by moving the left arm forward, as indicated by the gesture specification, but at the same time tried to grasp or press the visual “objects” (either menu items, oven functions, or recipes) on display. In addition, from our observations users tend in many cases to execute this gesture by tensing the arm and stretching it completely, which creates muscular stress and increases fatigue.

	Median	mean	stdev	RDS%
FUN	4	4,25	0,77	18,1%
INTUITIVE	4	3,78	0,83	21,8%
NATURAL	4	3,47	0,89	25,7%
IRRITATING	2	2,02	0,71	35,3%
INCOMPREHENSIBLE	2	1,85	0,64	34,8%
TIRING	2	2,16	0,83	38,5%

Table 1. User evaluation of the overall user experience

Finally, the last diagram pinpoints that the visual feedbacks are considered appropriate for over 85% of the users, confirming the quality of our visual interface design strategy.

Table 1 shows that the touchless gestural paradigm is largely perceived as fun, intuitive, and natural. The perception of fatigue is weak, even if the value of the relative standard deviation pinpoint that this perception is very subjective. Concerning *utility*, 82% of participants believe that touchless gestural interaction can be applied on contexts different from gaming (with domotics and tourism the most frequently mentioned application domains).

DISCUSSION AND CONCLUSIONS

There is a wide debate among interaction designers about touchless gestural interaction, its naturalness, and its role in modern applications [18][20][21][22]. Our work may provide a contribution to this discussion and increase our understanding on the design issues involved in touchless gestural interaction in-the-small. Touchless gestural interaction in-the-small represents a field where existing design guidelines of touchless gestural interaction in-the-large need to be validated, exemplified, and refined, in order to cope with the intrinsic limitations induced by the use of small displays, and the a strong *interplay* between visual design, gesture design, and ergonomics. The MOTIC example shows that the definition and implementation of gestures should be carefully considered in relationship with both the desired outcome of the application and the requirements or constraints set by both the size of the visual interface and the context of use. The latter include the need for a smooth, non-intrusive integration of the display with the physical environment, e.g., fitting the screen in the limited space of an oven in a cupboard.

MOTIC pinpoints that *touchless gestures can be a powerful mode of interaction*, even with small displays, although most gestures might not be “natural” in a strict sense, nor are easy to learn and to remember. Today, a touchless gestural language can hardly be defined to be the best one for any application and any user. Still, some authors [20][21] predict that, like it happened the mouse and the keyboard, gestures will become natural due to the widespread use of devices and applications adopting it, or by effect of standardization, either by a formal standards body or simply by convention, so that at some point of time the same gestures will be adopted and will mean the same things in different systems. “Natural” gestural languages need time to be better developed, for us to understand how best to deploy them, and for standard conventions to develop. Experiments and empirical validations like the one reported in this paper will help us to move in this direction. Today motion sensing and processing technology supports commercial, off-the-shelf, ready-made products (like Kinect) and enables low-cost touchless gesture based applications be developed, tested, and delivered relatively easily. This scenario is promoting the maturity of the field.

Finally, the MOTIC project is an example of how touchless gestural interaction can be used in conjunction with ordinary objects of our everyday activities and how it can play in multiple contexts of people’s life. From an application perspective, MOTIC is unique, to our knowledge. The marketing analysis carried on by Candy-Hoover shows that MOTIC represents the only fully implemented and tested application of touchless gestural interaction existing today in the house appliance domain, and can promote further exploitation of this paradigm in industrial contexts. For example, after the success at the Furniture Fair, Candy-

Hoover is considering not only to transform MOTIC from a promotional system to a commercial product, but also to apply its interaction paradigm to other household appliances produced by the company (e.g., washing machines). In addition, with other industrial partner we are exploring the possibility of using touchless gestural in-the-small interfaces for general house automation control.

REFERENCES

1. Ali Bigdelou, Loren Arthur Schwarz, N. N. An adaptive solution for intra-operative gesture-based human-machine interaction. Proc. IUI 2012, ACM (2012), 75-84
2. Ashbrook, D., Starner, T. Magic: a motion gesture design tool. Proc. CHI 2010, ACM (2010), 2159–2168.
3. Baudel, T., and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. CACM 36, 7 (1993), 28–35.
4. Bellucci, A., Malizia, A., Diaz, P., Aedo, I. Human-display interaction technology: Emerging remote interfaces for pervasive display environments. Pervasive Computing, IEEE 9, 2 (2010), 72 –76.
5. Benavidez, P., and Jamshidi, M. Mobile robot navigation and target tracking system. Proc. SoSE 2011, (2011), 299 –304.
6. Beyer, G., Alt, F., Müller, J., Schmidt, A., Isakovic, K., Klose, S., Schiewe, M., and Haulsen, I. Audience behavior around large interactive cylindrical screens. Proc. CHI 2011, ACM (2011), 1021–1030.
7. Cassell, J. A framework for gesture generation and interpretation. Computer Vision in Human-Machine Interaction (1998), 191–215.
8. De la Barré, R., Chojecki, P., Leiner, U., Mühlbach, L., and Ruschin, D. Touchless interaction-novel chances and challenges. In Human-Computer Interaction. Novel Interaction Methods and Techniques, J. Jacko, (ed.), LNCS 5611Springer (2009) 161–169.
9. Döring, T., Kern, D., Marshall, P., Pfeiffer, M., Schöning, S., Gruhn, V, Schmidt, A. Gestural interaction on the steering wheel: reducing the visual demand. Proc. CHI 2011, ACM (2011), 483-492
10. Grandhi, S. A., Joue, G., and Mittelberg, I. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. Proc. CHI 2011, ACM (2011), 821–824.
11. http://www.microsoft.com/en us/kinectforwindows/
12. http://www.xbox.com/en us/kinect
13. Johnson, R., O'Hara, K., Sellen, A., Cousins, C., and Criminisi, A. Exploring the potential for touchless interaction in image-guided interventional radiology. Proc. CHI 2011, ACM (2011), 3323–3332.
14. Heydekorn, J., Frisch, M., Dachselt, R.: Prospects of user elicited gestural interaction techniques. CHI 2010 Workshop on Natural User Interfaces: The Prospect and Challenge of Touch and Gestural Computing..
15. Kuikkaniemi, K., Jacucci, G., Turpeinen, M., Hoggan, E., and Muandller, J. From space to stage: How interactive screens will change urban life. Computer 44, 6 (2011), 40 –47.
16. Kurtenbach, G., and Hulteen. , E. Gestures in Human-Computer Communication. Addison-Wesley(1990).
17. Maged N., Kamel B., Bryan J., Blanchard, C. W. J. M. A. T., and Gutierrez-Osuna, R. Web gis in practice: a microsoft kinect natural user interface for google earth navigation. Int. J. health geographics (2011), 10-45.
18. Malizia, A. and Bellucci, A. The artificiality of natural user interfaces. CACM 55, 3 (March 2012), 36-38.
19. Ng, J., Sim, T.-J., Foo, Y.-S., and Yeo, V. Gesture-based interaction with virtual 3d objects on large display: what makes it fun? Proc. CHI EA 2009, ACM (2009), 3751–3756.
20. Nielsen, M., et al. Moeslund, T., Storring, M. and Granum, E. A Procedure for developing intuitive and ergonomic gesture interfaces for HCI. Proc. Gesture-Based Communication in Human-Computer Interaction: 5th Int. Gesture Workshop, GW 2003, LNCS 2915, Springer, Berlin (2004), 409-420.
21. Norman, D. A, Natural user interfaces are not natural. interactions, v.17 n.3, May + June 2010
22. Norman, D. A, Nielsen, J. Gestural interfaces: a step backward in usability. interactions, v.17 n.5, September + October 2010
23. Pickering, C. A, Burnham, K. J, Richardson, M.J. A Research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction. Proc. Automotive Electronics, 3rd Institution of Engineering and Technology Conference (2007), 1-15
24. Riener A. Gestural Interaction in Vehicular Applications. IEEE Computer, 42-47
25. Scoditti, A., Blanch, R., and Coutaz.J. A novel taxonomy for gestural interaction techniques based on accelerometers. Proc. IUI 2011, ACM (2011), 63-72
26. Shan, C. Gesture control for consumer electronics. In Multimedia Interaction and Intelligent User Interfaces, L. Shao, C. Shan, J. Luo, and M. Etoh, (eds.), Advances in Pattern Recognition. Springer, 2010, pp. 107–128.
27. Sven, G., Markus L., F. D. M. B. A. K. Using intelligent natural user interfaces to support sales conversations. Proc. IUI 2012, ACM (2012), 97-100
28. Vatavu, R.,. 2012. User-defined gestures for free-hand TV control.In Proceedings of the 10th European conference on Interactive tv and video (EuroiTV '12). ACM, 45-48.
29. Wachs, J. P., Kölsch, M., Stern, H., and Edan, Y. Vision-based hand-gesture applications. CACM 54, 2 (2011), 60–71.
30. Wei, J., Wang, X., Peiris, R. L., Choi, Y., Martinez, X. R., Tache, R., Koh, J. T. K. V., Halupka, V., and Cheok, A. D. Codine: an interactive multi-sensory system for remote dining. Proc. UbiComp 2011, ACM (2011), 21–30.
31. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. 2009. User-defined gestures for surface computing. In Proc. CHI '09. ACM (2009), 1083-1092.