Human-centric Computing
and Information Sciences

RESEARCH   **Open Access**

Proposal and testing goals-guided interaction for occasional users

Antonio L. Carrillo* and Juan A. Falgueras

*Correspondence:
alcarrillo@uma.es
Department of Computer
Sciences, University
of Málaga, International
Campus of Excellence
Andalucía TECH,
29071 Málaga, Teatinos, Spain

Abstract

The latest shifts in technology have brought about new kinds of users who occasionally access unfamiliar systems in new scenarios. This way of use should not request any learning curve. There have been many attempts to help this kind of users: agents, floating help, tooltips, direct video demonstrations, etc., elements that support the appealing direct manipulation style (DM), but add indeed an extra layer of supporting material over it. In this paper, we describe an analyse *goals-guided Interaction*, GGI, as the basis for a different approach. We do not expect the user to understand or use a metaphor but guide them throughout the whole interaction process. This help system is integrated in the interface. The user will be closely guided on the "what to do" and on the "how to do," always according to their current goal. This no-previous-knowledge-required way of interaction shortens significantly the learning curve. In order to support this proposal, a complete study based on user testing is performed. The results show that, for non-expert occasional users, GGI is more efficient in time and less error prone than the classic Direct Manipulation. Participants also showed a strong preference for GGI. This research offers an alternative way of interaction for designers and developers of human-centric interactive systems who have occasional users as their main target users.

Keywords: Interaction styles, Guided user interfaces, Occasional use, IoT, Deep learning, Adaptive interfaces, Human-Centric Computing

Introduction

One of the main research areas of *human–computer interaction* is the study of the different ways in which users communicate or interact with the computer [1, 2]. Each *interaction style* offers its own way of organising system functionality, managing user inputs, and displaying information. Two main approaches can be considered in order to interact with modern devices: the *conversational world* and the *model world*. The former is sequential and based on *text*. The latter, *the model world,* uses graphics and metaphors [3], like "Windows, Icons, Menus and Pointers" (WIMP), to assist the user with an asynchronous and a free management of objects on the screen. Users can see and predict the behaviour of familiar objects through metaphors. They then follow their natural intuition to manipulate them, receiving immediate feedback. The success of this *Direct*

Manipulation (DM) [1, 4, 5] interaction style has broadly eclipsed others, and these interfaces are typically assumed to be the best solution for every possible scenario, task and user.

However, the massive expansion of technology has caused the appearance of new scenarios of use in addition to quite different tasks and users for them. This requires revisiting the most appropriate way of interaction for each case, because the success of the interaction depends basically upon its suitability for the final user. According to Coe [6], there are many differences between how *experts* and *novices* perceive and use software applications. On the one hand, *experts* have the best possible mental model thanks to their experience. As opposed to *novices*, they require less guidance and help. On the other hand, *novices* are users with little or no knowledge about the system. They initially have a deficient understanding about the possibilities of the application and have to focus on how to deal with the interface [7]. To gain expertise they follow a *learning curve* that is usually traversed through study, trial and error, and repeated use of the application [8, 9]. However, what happens when the user needs to use the system in an occasional way, and does not want or cannot afford a learning curve? What if the user needs to use the system with an irregular frequency, and whatever is remembered from previous uses, if any, does not provide them with enough knowledge for the current use?

We are addressing the problems that arise when a system is occasionally accessed. At least three arguments support an explicit design of systems for these *occasional* users. Firstly, the evolution of *Graphical User Interface* (GUI) [1] and the correct inclusion of *multi-touch* systems [10–12] have brought about new kinds of devices and new ways of interaction. Secondly, this increase in the heterogeneity and number of occasional users is also driven by the incorporation of the principles of *Usability* [13–15], *Inclusive Design* [16, 17], and *Accessibility* [18–21]. Thirdly, the ubiquity and the permanent *online* interconnection of the *devices* have burst into less traditional tasks and contexts of use, e.g. shopping centres, leisure facilities, museums, airports… [22–24]. These technologies have facilitated new scenarios of use where human–computer interaction is on the move, using technology as a mean to accomplish an immediate goal.

These occasional scenarios and users demand simplicity, immediacy and, ideally, no prior knowledge about the use of the system. This means that even if the user dealt previously with the same or analogous interface, there are a number of factors that make it unwise to rely on the user's memory recall or implicit visual recognition as the main mechanisms to learn how to use the system. Some of those factors could be the time elapsed since the last interaction, the difficulties users experience on learning, and, in many cases, their lack of interest [25]. It is recommended to assume that the user will have to cope with an unknown interface, with no requirement upon previous knowledge to use it.

In this paper, we present and test *goals-guided interaction (GGI)*, that is an alternative way of interaction especially suitable for occasional users and, in a more general way, for those users who do not want or cannot afford a learning curve. Elderly novice users [26] are also good candidates for this interaction style. Now, in the next section, we will describe GGI in more detail. However, our main aim is to check with real users what kind of interaction they prefer in these occasional situations. Therefore, "An empirical comparison between goals-guided interaction and conventional direct manipulation"

section presents an empirical study comparing two versions of the same application for furnishing kitchens, one with a *GGI interface* and the other with a conventional *DM interface.* In "Discussion" section we discuss the results and, finally, in "Conclusions" section we draw conclusions.

Goals-Guided Interaction

Goals-Guided Interaction (GGI) is an interaction style that pursues the ease of use without relying on the traditional mechanisms of learning, like repetition and retention. GGI does not expect the occasional user to understand a metaphor. The user does not need to be familiar with graphical widgets or think what the next command should be. No previous knowledge is required, shortening significantly the learning curve, if any.

In order to achieve these challenging purposes, this interaction style guides the user along the whole interaction process, step by step, in a hierarchical and progressive way, according to their current goal. It is conversational and sequential in nature. It does not allow performing tasks in parallel as pure DM interfaces could. On the other hand, this does not prevent the user from deciding among alternatives. As will be explain in "Fundamentals and aims of GGI and summary of the methodology and notation for the specification of GGI interfaces" sections, hierarchically organised goals and sub-goals will be explicitly presented to the user, one at a time. These objectives, once achieved, are left behind. Nevertheless, the user will have the possibility to return to previous goals, via cancellation, whenever it is possible to do so.

The user will be strictly guided on the "what to do" and the "how to do it," one step at a time, by means of a guide or help system integrated in the interface. This strategy constrains the freedom of the DM in a similar way to processes guided by *wizards* [27–29]. In some sense, GGI could be viewed as a layer over pure DM, making it a kind of *assisted interaction*, adding control over DM: it allows the user to touch, move or interact with graphic objects, but only when the guide allows for it. It could also be considered as a super-set of wizards. Standard DM transfers the control to the user whereas GGI guides them.

Related work

Giraud et al. [21] conducts relevant experiments showing that filtering redundant and irrelevant information for new user with special needs is paramount for their accessibility. Our proposal, *Goal-Guided Interaction* (GGI), has as main goal guiding the user, avoiding they needing to deal with metaphors and irrelevant information for the goals they want to accomplish. GGI exposes them only to the steps they have to take according to their current goal.

Jo and Kim [30] describe a scenario which uses an Internet of Thinks (IoT) and Augmented Reality (AR) service on a smart pad based on a step-by-step how-to-use guideline. The user wants to print on an unknown printer first time and does not even have a user manual. In this case, advanced technology could help quite a lot, but the underground situation is similar to those GGI affords: guiding the user, step by step, on the "what to do" and the "how to do it". Our proposal is simpler to implement in any device. Of course, using AR could be a bonus.

The idea that "unsolicited advice helped performance more than advice requested by the advisee (the person who asks for help), but only for unfamiliar tasks and especially for difficult tasks" is well explained in [31]. On the other hand, non-expert users cannot rely on disciplined training or previous familiarity with systems especially with the handle of security and the prevention of hacking, phishing, etc., It can be an impossible challenge for inexperienced users to anticipate the vulnerabilities in the security of systems [32, 33].

Gillingham [34] shows that for real welfare agencies taking data from children, it would be more efficient to use simple guided systems than complex systems based on direct manipulation. Compared to complex information systems (IS), simple IS provide managers with limited control over what practitioners do, how they can do it, and when, so guiding practitioners limits the number of errors they can make. This is another objective of GGI: to offer decision makers the possibility enumeration of the specific steps the user must follow, reducing the complexity of the system and the possibility of making mistakes.

Holl et al. [35] develops an on-site mobile evaluation approach that allows to evaluate user interaction that was studied during the FIFA World Cup 2014. For the evaluation of the app, they faced the problem of participants in a large event, such as a soccer match, were not willing to spend time on completing a long survey or interview. People were experiencing an emergency situation having their cognitive capabilities affected by emotional burden, so mobile apps should be the easiest and the most intuitive to interact with. GGI guides the user thorough steps not demanding much attention from them in emergency or occasional situations.

Williams and Berlanstein [29] decided to create guiding wizards for the most difficult tasks of a library website. Finally, they could verify that it was a success. Those wizards empowered users to find answers to their questions virtually, especially during non-business hours. The wizards were interactive forms that provided choices, and hided and revealed content, based on users' answers. These solutions also share some characteristics with the philosophy of the GGI.

Fundamentals and aims of GGI

The fundamentals of *Goal-Guided Interaction* are based on the mechanisms of human reasoning for *problem resolution*, breaking up the main goal into a hierarchical tree of sub-goals. The leaves of the tree are the elementary actions or the final sub-goals not requiring further explanations or decomposition. We also conceive the interaction as a *problem resolution task* and consider the user as an information processing system. This vision lets us describe all the activities that take part in the interaction process in a similar way as GOMS technique does [36, 37]. A GOMS (Goals, Operators, Methods, and Selection rules) model is a description of the knowledge a user must have in order to carry out tasks and, thereby, accomplish goals; it is a representation of the "how to do it", the steps and decisions to take. The aim of GGI is to preclude the user from having to devote time to acquire that knowledge, by integrating it as a guide in the interface.

For this purpose, it is fundamental to perform a task analysis [38, 39] and the specification of the interface. NGOMSL (Natural GOMS Language) [40, 41] proposed by Kieras is the most widespread and studied variant of GOMS and the closest to the
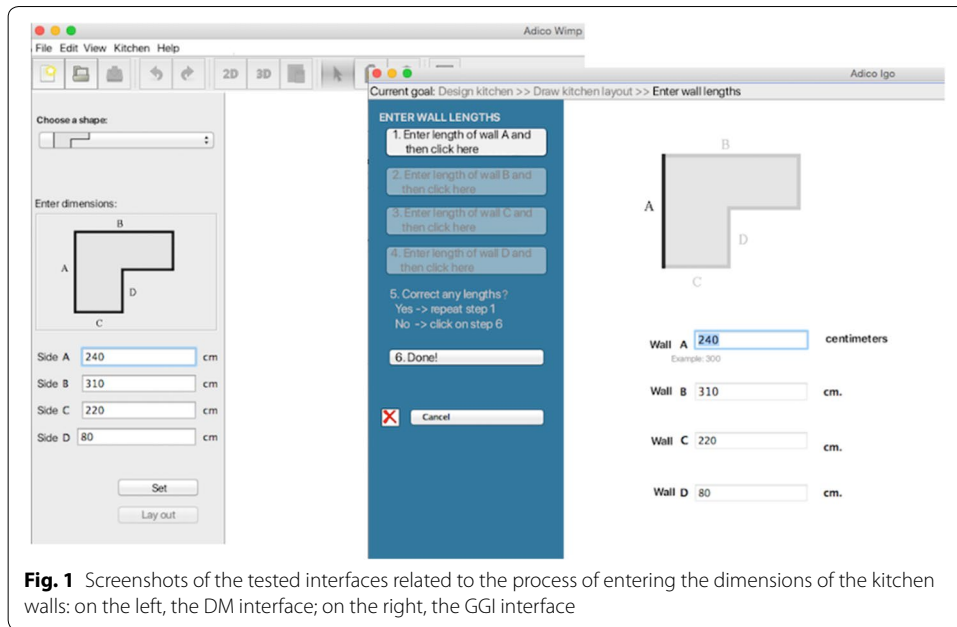
Carrillo and Falgueras *Hum. Cent. Comput. Inf. Sci.*   (2020) 10:4

Page 5 of 21

Fig. 1 Screenshots of the tested interfaces related to the process of entering the dimensions of the kitchen walls: on the left, the DM interface; on the right, the GGI interface
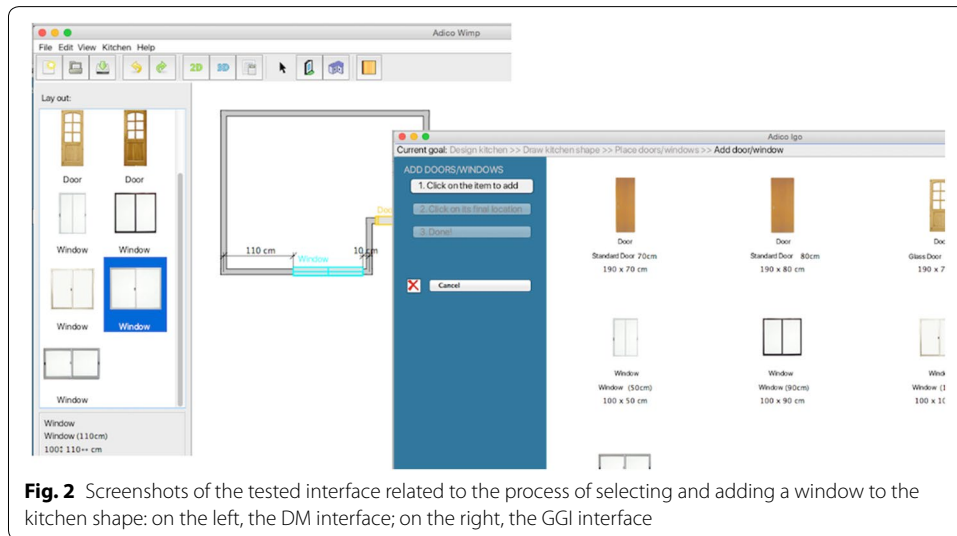

Fig. 2 Screenshots of the tested interface related to the process of selecting and adding a window to the kitchen shape: on the left, the DM interface; on the right, the GGI interface

user's natural language: any user can read and follow it, as a cooking recipe. We have adapted and extended NGOMSL to use it as the basis for the specification of the guiding part of GGI interfaces. A specification made with our extended NGOMSL, after an easy compilation process, can be used to build the *Goal Guided Area* (*GGA*) of the GGI *interface.*

GGA is what explicitly guides the user through the hierarchy of goals. This area, that we placed on the left (and with a blue background) of the GGI interface shown in Figs. 1, 2 and 3, becomes the substitute for the typical menus and toolbars used in standard DM interfaces, which are not necessary for these guided interfaces. GGA is the most important part of GGI interfaces. It is always present, and always shows the
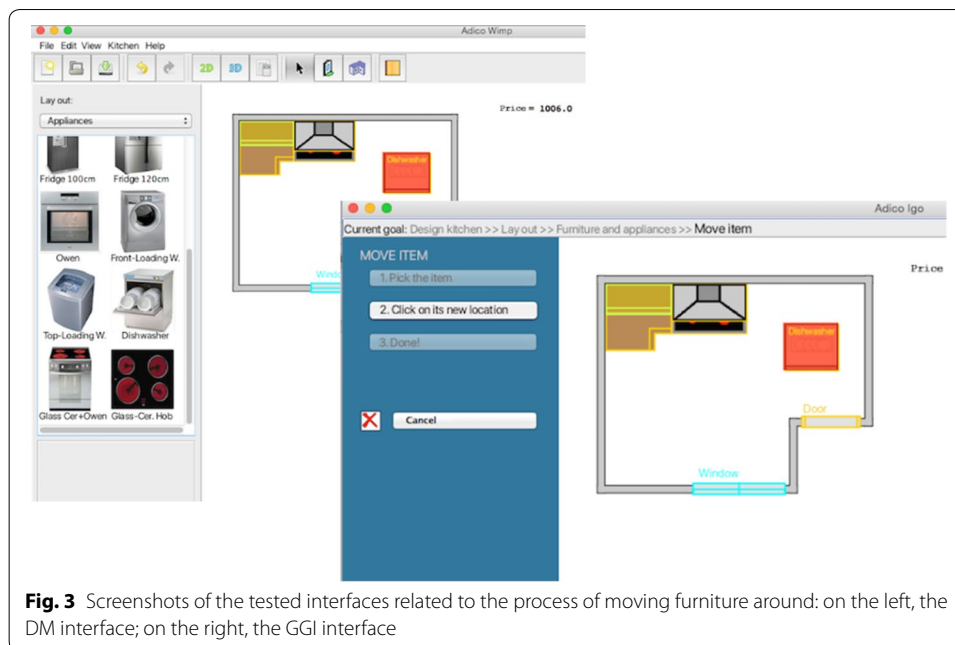
Carrillo and Falgueras *Hum. Cent. Comput. Inf. Sci.* (2020) 10:4

Page 6 of 21

Fig. 3 Screenshots of the tested interfaces related to the process of moving furniture around: on the left, the DM interface; on the right, the GGI interface

method or the *selection* that allows the user to accomplish the current *goal*. If this *goal* requires the user to make a *selection*, *GGA can* show the different excluding *options* (*alternatives*) that compose that *selection*, so that the user can choose one of them. In the case of the *goal* must be accomplished following a specific *method*, *GGA* will offer the sequence of *steps* that compose this *method*. The interface underlines the current *step, which* would imply the launching of a new *sub-goal* (and then, another *method* or *selection*), or the realisation of an elementary *action* (without *methods* nor *selections*).

Recently, Gallagher et al. [42] conducted an interesting study to compare the performance and impact of multiple computer monitor with the use of a single monitor. There was a strong evidence that users prefer dual-monitor displays. They also measured an increase in task efficiency. The dual-area display structure of a GGI interface (the GGA and the Working Area) could be one of the most appropriated applications of those configurations.

Summary of the methodology and notation for the specification of GGI interfaces

As indicated above, the recommended methodology and notation for the specification of a GGI interface are based on and expand NGOMSL. In this sense, the specification of these interfaces is based on *goals*, *actions*, *methods*, and *selections*. *Goals* are intentions the user has. *Actions*, *operators* in NGOMSL, are simple cognitive, perceptive or motor activities. They could be an elementary activity, such as hitting a key, or a high-level one the users are able to carry out by themselves, no needing more details. A *method*, as in NGMOMSL, is the sequence of *steps* (or *actions*) the user needs to follow to accomplish a *goal*. And *selections* play a role similar to the *selection rules* in NGOMSL. A *selection* can be built using as many mutually excluding *options* as necessary. Each *option* is associated with a condition the user has to check. After they have decided, they must start

the associated sub-*goal*, which, again, will be performed by another *method* or another *selection.*

Moreover, GGI notation expands the basic NGOMSL constructors, adding four new descriptors that provide more dynamicity to the models: *conditional* steps/options, steps/options with *effects, cancelability,* and *selections for the system*:

- In the case of *conditional* steps/options, the system, while running, must check the internal states specified by the corresponding clause, enabling or showing to the user, or disabling or hiding from the user, those *steps/options.*
- Each step/option could be appended with an effect clause, followed by the description of that effect on the system.
- Another clause offers to the user the opportunity to *cancel* the current *goal* and returns to the father goal.
- In general, *selections* are meant for the user to make. Nevertheless, sometimes, during the interaction process, we want the user to see new added goals, depending on internal states that the system has to check. They are *selections for the system*.

The details of the GGI methodology and notation (Additional file 1), and the specification of the tested GGI interface can be consulted in the Additional file 2 (https://github.com/juanfal/GGI).

An empirical comparison between Goals-Guided Interaction and conventional Direct Manipulation

In previous sections, we explained that GGI is proposed as an alternative way of interaction to help occasional users who do not want or cannot afford a learning curve for specific goals. Now, we want to test this postulate with real users, conducting a comparative study between guided (GGI) and non-guided (DM) interaction. This empirical analysis will give us objective measurements and subjective opinions. We want to test 2 objectively measurable hypotheses, related to time on task and errors, and a third subjective hypothesis, related to satisfaction.

For occasional non-computer-expert users:

- (H1) they proceed quicker with GGI than with standard DM,
- (H2) they make fewer errors with GGI than with standard DM,
- (H3) they prefer GGI to standard DM.

As can be seen, the 3 pillars of usability are covered: efficiency (H1), effectivity (H2), and satisfaction (H3).

Participants
The study involved twenty volunteer participants (n = 20), a number that meets the criteria of [43–45]. There were 12 females and 8 males. Their ages ranged from 12 to 52 years, with a mean age of 44 years. All subjects were unaware of the final aim of the

research. They had not participated in previous usability studies nor had they received any incentives to participate in the experiment.

Prior to the main study, a pilot experiment was carried out. It involved a group of 4 participants. This experiment allowed us to fine-tune the test procedure, the time span, and the descriptions of each task and questionnaire.

Table 1 summarises the main data and characteristics of the participants, collected by means of a *background questionnaire.*

To do the recruiting and screening of the participants, the primary requirement was *no previous experience* with similar applications to those tested in this study. Only two exceptions were intentionally included: two participants were professionals at the task domain. Both had this task as part of their daily activity using similar commercial direct manipulation software packages. The other screening criterion was to discard *computer* expert users. As reflected in Table 1, no participant had a self-graded

Table 1 Overview of the main data and characteristics of the participants, collected by means of the *background questionnaire*: *gender* (G), *age* (A), *computer expertise* (CE), *previously used software* (PUS) [options: Internet, Email, Word Processor, Spread_sheet, Multimedia, Accounting & management, Databases, social Networks, Others], *Kitchen Design Software Knowledge* (KDSK), *Tablets and Smartphones Expertise* (TSE), *Task Domain Knowledge* (TDK), *Academic Studies Level* (ASL) ranged from 0 to 3 [0: No studies; 1: Primary school; 2: Secondary school; 3: University], and *Other Relevant Knowledge* (ORK). The final CE of each participant was confirmed with all these items

#	G	A	CE	PUS	KDSK	TSE	TDK	ASL	ORK
	M/F		[0..5]	[I, E, W, S, M, A, D, N ,O]		[0..4]	[0..4]	[0..3]	
1	F	51	0		No	1	3	1	
2	F	48	0		No	1	3	1	
3	F	48	0	I	No	1	3	1	
4	M	12	0	I	No	2	1	0	
5	F	42	1	I, E	No	2	3	1	
6	F	50	1	I, M	No	2	3	2	
7	M	24	1	I, M, O	No	2	2	1	Occasional use of IKEA kitchen design app
8	F	46	2	I, E, W, S, M, A (forgotten)	No	2	3	2	Forgotten computer courses
9	F	44	2	I, E, W, M, N, O	No	2	3	1	
10	F	50	2	I, E, W, M, O	No	2	3	1	Medical software
11	F	40	2	I, E, W, M, O	No	2	3	3	Medical software
12	M	52	2	I, E, W, S, M, O (forgotten)	No	2	3	1	Pantograph. forgotten computer courses
13	M	43	2	I, E, W, M, A, N, O	No	2	2	1	Photoshop
14	F	35	3	I, E, W, S, M, A, D	No	2	3	2	
15	M	43	3	I, E, W, S, M, A, D, O	No	2	2	3	Veterinary software
16	M	50	3	I, E, W, M, O	No	2	3	1	AutoCAD and Presto user
17	F	43	3	I, E, W, S, M, A, D, O	No	1	3	2	Forgotten computer courses
18	M	52	3	I, E, W, S, M, A, D, O	No	2	3	2	
19	M	52	3	I, E, W, S, M, A, O	Kitchens	3	4	2	Photoshop Kitchen design professional
20	F	52	3	I, E, W, S, M, O	Kitchens	2	4	1	Kitchen design professional

Computer Expertise (CE) greater than 3 on a 5–point scale. This parameter is subjective. To confirm these self-assigned values, we also asked several complementary questions: PUS, KDSK, TSE, TDK, ASL, and ORK (see Table 1).

Aside from the main study, we conducted a quick and simple eye tracking experiment only for two participants randomly chosen. See "Results from a short eye tracker analysis" section for details.

Materials and tasks

We developed two versions of the same application (in Java language), one with a conventional *DM interface* and the other with a *GGI interface* (Fig. 1). Both share most of the code except for the sections in which the user interfaces are involved. Both offer the same functionality and allow the user to achieve the same goals. The DM version is inspired by the desktop version offered by IKEA [46].

Although other applications could be plausible, we chose for our study a familiar domain, i.e. kitchen furniture. This scenario comprises three development stages: designing, furnishing and final adjustments of a kitchen. Some of the involved tasks require extensive object manipulation, usually implemented with standard DM interfaces. These characteristics could lead us to think that this is an application especially suitable for a classic DM interaction style, making it particularly challenging for our study.

Figures 1, 2 and 3 depict sub-tasks for some of the stages the user goes through. The screenshots on the left correspond to the DM interface whereas the screenshots of the right are related to the alternative GGI interface. The DM interface is a plain standard DM-WIMP application. Firstly, the user introduces the shape and the dimensions of the kitchen and then furnishes it. To this end, the user has to pick elements and position them inside the boundaries following the canonical DM way, i.e. with undo/redo, tooltips for buttons, right-click pop-up menus, etc.

Regarding the three screenshots corresponding to the GGI interface, note that the user is presented with a list of actions to be read and followed. Each action is atomically performed by short DM interactions in the Working Area. The user has to finish the current step before going to the next, as was explained in "Goals-guided interaction" section.

- A screencast of a participant using GDI can be watched in https://youtu.be/4owtCVMuqc4.
- A screencast of a participant using DM is shown in https://youtu.be/QeafXLlsXzY.
- The source code for both applications is available in https://github.com/juanfal/ADICO.

Each participant was given a single piece of paper containing the information shown below. It briefly described the application scenario along with the three groups of tasks chosen for the test.

Imagine yourself arriving at a kitchen furniture shop. All employees are busy. However, a computer is available for you to specify your preferences for your kitchen: shape, dimensions, the furniture you want to buy and where you would particularly like to place it, make further adjustments, etc. To save time, you decide to use this system

Task 1: Specifying how your kitchen is
Suppose that you bring a sketch (as the figure below) with the shape of your kitchen, the dimensions of the walls, doors and windows, and the location of all of them:
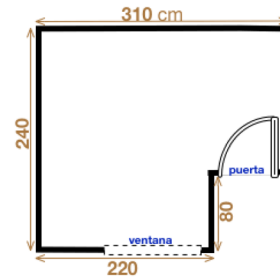
- Door 190 (high) × 70 (width) cm
- Window 100 (high) x 110 (width) cm

Task 2: Choosing furniture
Suppose the furniture you want to buy is listed below
Choose and place the pieces as shown in the figure below:
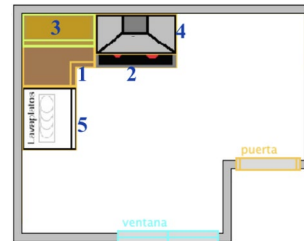
1. Kitchen corner (80 cm wide)
2. Hob with oven included
3. Showcase (80 cm, with 2 doors)
4. Metal extractor hood
5. Dishwasher

Write down the budget: _____ (euros)

Task 3: Retouching and ordering
- Place the dishwasher next to the hob.
- A family is going to give you a metal extractor hood, therefore remove ours
- Observe the kitchen in 3D
- Look and see the detailed kitchen budget and note down the number of chosen items: __and the final budget: __ (euros)
- If you think there is something important to be done (before commissioning the kitchen), do so
- Commission the kitchen (to confirm with a kitchen seller)

Study design

Our study used a *Within-Subjects Design* with one independent variable, namely the type of *interaction*. This independent variable has two levels: a GGI interaction and a classic DM interaction. This *Within-Subjects Design* or *Repeated Measures Design* was necessary since we wanted each participant experienced both styles of interaction. It was important that both treatment groups exactly included the same participants. We wanted them to compare both interfaces expressing their preferences, and not to worry about personal characteristics that could bias our results.

On the other hand, in order to mitigate the potential transfer-of-learning effect caused by testing one type of interface prior to the other, a *counterbalanced* strategy was applied alternating the order among users. Half of the participants started with GGI, while the other half started with DM.

Measurements

Quantitative measurements: time on tasks and number of errors

The first empirical data to be considered in this study is *time on task*. In this sense, we will denote as T_{task1}, T_{task2} and T_{task3}, the time, in seconds, that participants spent on each of the three groups of tasks, respectively. These tasks were labelled in the task sheet ("Materials and tasks" section) as "Task 1: Specifying how your kitchen is," "Task 2: Choosing furniture" and "Task 3: Retouching and ordering," respectively. T_{total} will be the total time, i.e. the sum of T_{task1}, T_{task2}, T_{task3}.

The analysis of *types of errors* is an important and necessary complement to the timing analysis. If severe errors occurred (without the assistance of the moderator), then the time on task would be highly affected, or even worse, could prevent the user from completing the task.

Three degrees of *errors* have been considered, namely *slight*, *moderate*, and *severe*:

- The *severe or* blocking error took place when the user was stuck with something in the interface, preventing them from finishing the task. The moderator always assisted users with severe mistakes in order to have them finishing all the tasks and collect the corresponding user data.
- *Moderate* errors corresponded to non-blocking mistakes– not necessarily detected by the user–that could alter the outcome. In some cases, these issues could require warnings from the moderator.
- Finally, *slight* errors could not be categorised as actual errors because the user could resolve them without any moderator assistance, yet perhaps spending a bit more than the typically expected time.

Table 2 displays the concrete list of *errors* detected during the tests including, among other things, the type of interface and the task in which they occurred.

Subjective measurements

We present two sets of subjective data in this paper, collected from two types of questionnaires. The first set is composed of two identical *post-test questionnaires* that participants filled immediately after finishing with each interface. These *post-test questionnaires* ask the participant seven questions (Q_i). All Q_i questions, except for Q_5, are scored with a numerical value. These questions are:

- Q_1: "Do you consider that the application has helped you in knowing *what to do* in each moment?".
- Q_2: "Do you consider that the application has helped you in knowing *how* to do it? (what you needed to do)".

Carrillo and Falgueras *Hum. Cent. Comput. Inf. Sci.* (2020) 10:4

Page 12 of 21

Table 2 Overview of the *errors* (*Slight*/*Moderate*/*Severe*) specifying the interface (DM/GGI) and the tasks involved

Type	Interface	#Task	Incidence description	Effect	Assistance
Slight	DM	1	UNDO was confused with CLEAR when wanting to remove an item like a door (the user thought the system had failed because the item was not deleted)		
	DM	1	The user accepted the default shape of the room; later he was aware of it and changed it		
	DM	1	Window dimensions were confused with its distance to the walls	Put the proper window by chance	
	DM	2 and 3	Tried to change the position of a door/window, while another operation (the placement of another door/window) was being in progress		
Moderate	GGI	1	Not consciously placing a window, believing it did not matter	Consciously leaves an incomplete task (though not decisive for a correct final result)	
	GGI	2	Instead of moving a piece of furniture (because this option was not found), deleted it and put another one at the targeted location	Inadequate procedure	
	GGI	3	The user could not see the kitchen in 3D	Consciously leaves an incomplete task (though not decisive for a correct final result)	
	DM	1	Entered the measurements in the default kitchen shape (without choose the correct shape), and did not realize below	Wrong final result, unconsciously	Warning
	DM	1	The participant was unable to change the shape of the default kitchen, so he decided to start again	Inadequate procedure	
	DM	2 and 3	UNDO was confused with CLEAR (wanting to remove a cabinet), throughout the experiment	Persistent misinterpretation of an interaction (Inadequate procedure)	
	DM	2 and 3	Learnt not to DRAG to move furniture → choose delete and put it back at the targeted location	Inadequate procedure	
	DM	3	Chosen wrong furniture (high corner instead of low)	Wrong final result, unconsciously	Warning
	DM	3	Oblivious to the only sub-task is not explicitly stated: "choose the common elements (granite, wood and handles)"	Wrong final result, unconsciously	Warning
Severe	GGI	3	With everything done, the user would choose "design a plan" (which implies starting again) instead of the "order kitchen" option, so intend to start again	It would enlarge the experiment time without limit	Help
	DM	2	Does not learn to SCROLL (necessary to choose the desired furniture)	Blocking	Help

Table 2 (continued)

Type	Interface	#Task	Incidence description	Effect	Assistance
	DM	2	Does not learn to DRAG (needed for "placing" the furniture)	Blocking	Help
	DM	2 and 3	Does not learn to DELETE furniture	Blocking	Help
	DM	3	Does not learn to DRAG (necessary to relocate furniture)	Blocking	Help
	DM	3	Does not learn to DELETE furniture → Try to start again	It would enlarge the experiment time without limit	Help

- Q_3: "What was *harder?* (1) knowing *what* to do, (2) knowing *how* to get it done, (0) both easy, or (3) both hard". The scores were chosen according to a range between the best and the worst-case scenarios.
- Q_4: "Would you like to have had a more complete *help* system inside the application interface?".
- Q_5: "Choose (multiple choice) types of *periodicity* of use you consider the application is appropriate for: Just once, very rarely (once a year), Once a month, daily".
- Q_6: "Would you *use* a similar application for the design of your next kitchen?".
- Q_7: "To summarise, grade how *easy* the application is *to use*".

The second set of data comes from one unique and final *comparative questionnaire* that participants filled at the very end, once they had used both interfaces. This questionnaire makes them directly *compare* both interaction styles through the next eight questions (C_i):

- C_1: "With which interface is it easier to know *what* to do in each moment?"
- C_2: "With which interface is it easier to know *how* to do it?"
- C_3: "Which interface should include *more help* systems?"
- C_4: "Which interface is *easier to use* and requires *less training*?"
- C_5: "Which interface lets you work *quicker*?"
- C_6: "Which interface would you recommend to a *computer professional* for an *occasional use*?"
- C_7: "Which interface would you recommend to a *kitchen design professional* for a *daily use*?"
- C_8: "Which interface would you *choose* for furnishing your kitchen?"

Study procedure

The estimated time per participant (including questionnaires) ranged from 45 to 75 min. The tests were performed individually in an interference-free environment, except for the presence of the moderator. A common laptop computer was used for all the tests. This facilitated their relocation and was less intimidating and more familiar to users than any other bigger equipment. Mouse and keyboard were the input devices. The whole

process of interaction was recorded from both the computer screen and the device microphone for further analysis when necessary.

The procedure for the study followed the next steps:

1. All participants signed an *informed consent* and were briefed about the fact that the tests were taken voluntarily, being free to leave at any time without any justification.
2. The users read the *moderator script* to ensure that each participant received the same information about the type of test, the purpose of the study, and the testing process. This document also clarified that these tests were not intended to make any personal or psychological assessment, but were exclusively conducted for the evaluation of the involved computer interfaces. Then, the moderator answered any participant questions or concerns, if any.
3. The participants filled out a preliminary *background questionnaire* in order to collect user characteristics, such as age, gender, computer skills, previous software knowledge, task domain knowledge, etc. This helped to check whether the participants met the screening criteria (described in "Participants" section).
4. The users received a single-sheet document briefly describing the three groups of tasks to perform (detailed in "Materials and tasks" section).
5. As mentioned above, in order to implement a *counterbalanced* strategy, each participant was asked to use both versions of the application (alternating the order after each user). The so-called *Thinking Aloud* technique was also applied, but in a relaxed way. The participants were not forced to explain what they were doing any time, but they were free to express themselves. The moderator tried either not to interfere at all, or to intervene only in some non-blocking situations and always in the few critical or blocking situations that took place. Every error was systematically noted down and categorised as part of the data collection. For every user and for each type of interface, the time the participant spent on achieving each task was written down as well. The whole process of interaction was recorded with a computer screen and voice recording software for further re-examination when necessary.
6. At the end of the use of each interface, but before moving on to the other, the participants filled out the corresponding *post-test questionnaire* (detailed in "Subjective measurements" section) on subjective and specific usability points about the interaction they have just experienced. This questionnaire covered important aspects for the final evaluation.
7. Finally, the participants filled out the *comparative questionnaire* (detailed in "Subjective measurements" section) for comparison of both styles of interaction, requiring them to directly express their preferences.

Results

On the one hand, in our experiment, we have *repeated measures designs* with two conditions, because *the same* participant uses both types of interfaces. On the other hand, to verify the significance of the results, the most appropriate statistical method is a non-parametric *Wilcoxon paired-sample test,* also known as *Wilcoxon signed-rank test* [47],

Carrillo and Falgueras *Hum. Cent. Comput. Inf. Sci.* (2020) 10:4

Page 15 of 21

because the data are not normally distributed and then do not meet the requirements for the corresponding parametric test, the paired sample *t* test.

We next present the results in the following order: first, in "Quantitative empirical measurements: time on task and number of errors" section, the quantitative performance measurements through the time on tasks and the number of errors; second, in "Subjective results from post-tests questionnaires" section, the *subjective* data about each interface; and, third, in "Results from the comparative questionnaire: the user preference" section, the personal preferences. Finally, in "Results from a short eye tracker analysis" section, we elaborate on the *Eye tracker* experiment.
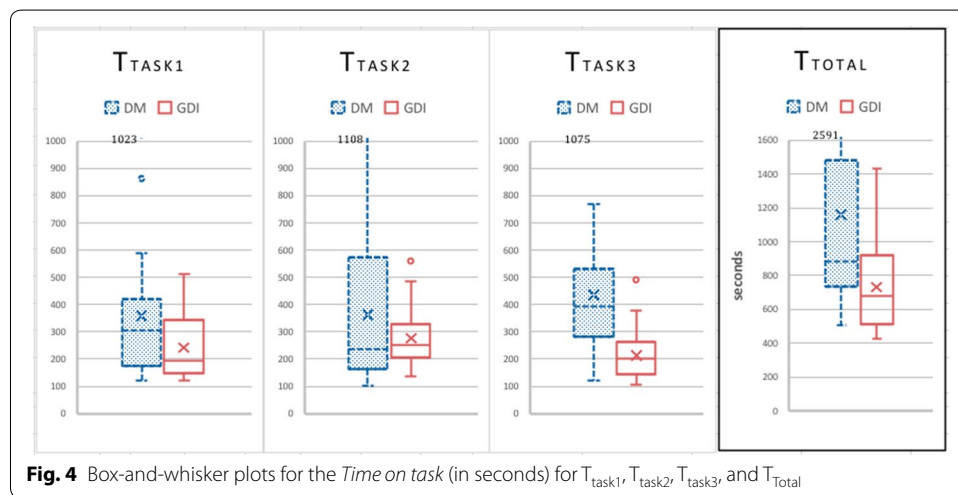

Fig. 4 Box-and-whisker plots for the *Time on task* (in seconds) for T_{task1}, T_{task2}, T_{task3}, and T_{Total}
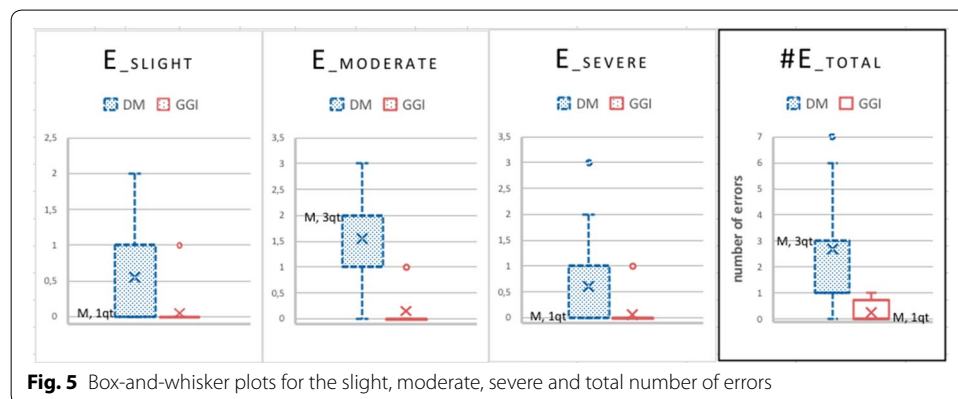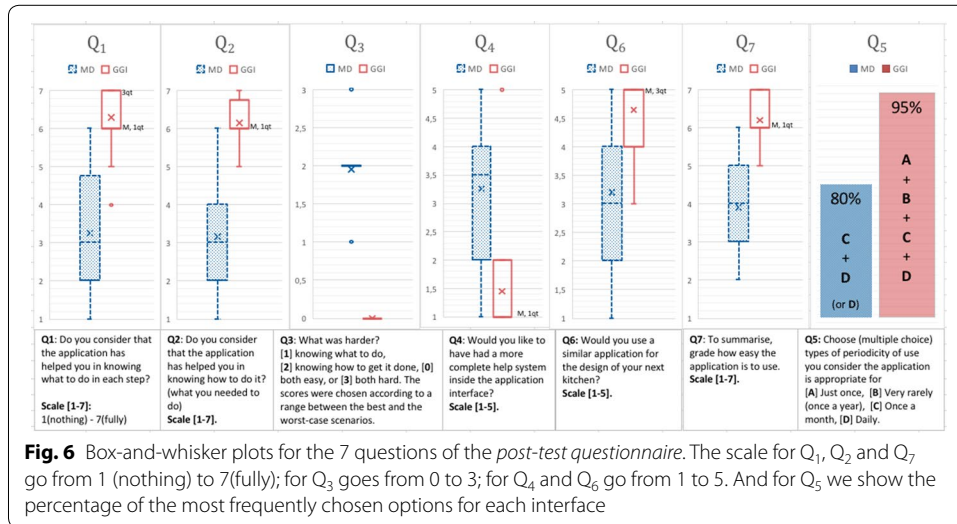

Fig. 5 Box-and-whisker plots for the slight, moderate, severe and total number of errors

Table 3 Results of the Wilcoxon tests on *time on tasks* and *number of errors*

Time on task	p-value	W	Type of test	#Errors	p-value	W	Type of test
T_{task1}	0.00250	180.5	One-tailed	E_{slight}	0.00750	50.5	One-tailed
T_{task2}	0.29430	134.0	Two-tailed	$E_{moderate}$	0.00019	136.0	One-tailed
T_{task3}	0.00007	207.5	One-tailed	E_{severe}	0.01800	33.0	One-tailed
T_{Total}	0.00001	204.0	One-tailed	E_{Total}	0.00009	171.0	One-tailed

Table 4 Some relevant results related to *time on tasks* and *number of errors*

Interaction style	% Earlier finishing	% Errors		
		Slight	Moderate	Severe
DM	5	92	91	92
GGI	95	8	9	8

Fig. 6 Box-and-whisker plots for the 7 questions of the *post-test questionnaire*. The scale for Q_1, Q_2 and Q_7 go from 1 (nothing) to 7(fully); for Q_3 goes from 0 to 3; for Q_4 and Q_6 go from 1 to 5. And for Q_5 we show the percentage of the most frequently chosen options for each interface

All the data tables (and the Additional file 3) can be downloaded from https://github.com/juanfal/GGI.

Quantitative empirical measurements: time on task and number of errors

Figures 4 and 5 display the box-and-whisker plots for the *time on task* (T_{task1}, T_{task2}, T_{task3}, and T_T) and for the number of *errors* (slight, moderate, and severe), respectively.
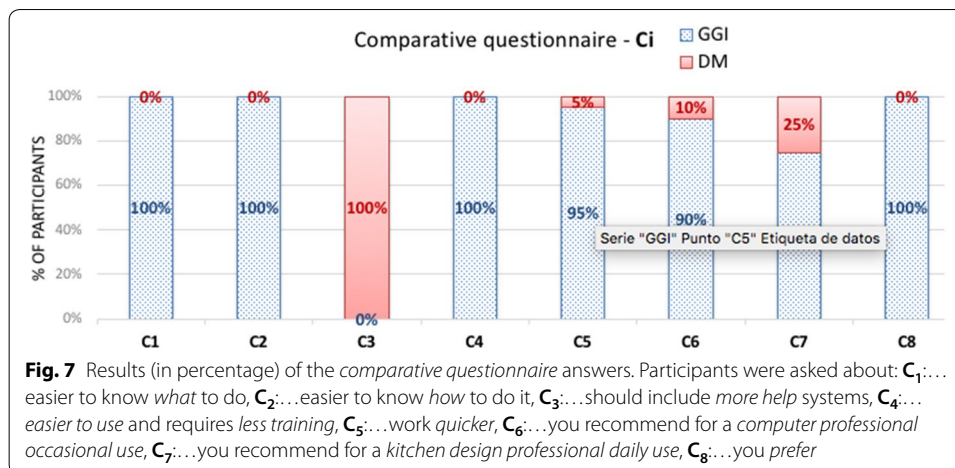
Concerning errors, the results of the *Wilcoxon* tests (Table 3) for the number of slight, moderate and severe errors produce significant differences ($p < 0.02$) in favour of GGI: DM presents a significant increase in the number of errors when compared with GGI. Therefore, we can accept H2.

Table 4 highlights how evident these results are by pointing out some relevant percentages: first, in accordance with T_T, 95% of users finished earlier using GGI than using the DM interface; second, 91% of the errors occurred under DM interaction versus only 9% under GGI.

Subjective results from post-tests questionnaires

In "Subjective measurements" section we remarked that the subjective *post-test questionnaire* used in this experiment included seven questions (Q_i) for the participants. They answered them as soon as they finished the tasks with each type of interface.

Figure 6 shows the box plot for the *post-test questionnaire*, except for Q_5 (not answered with a numerical value). According to the corresponding *Wilcoxon paired-sample tests*,

Fig. 7 Results (in percentage) of the *comparative questionnaire* answers. Participants were asked about: C_1:… easier to know *what* to do, C_2:…easier to know *how* to do it, C_3:…should include *more help* systems, C_4:… *easier to use* and requires *less training*, C_5:…work *quicker*, C_6:…you recommend for a *computer professional occasional use*, C_7:…you recommend for a *kitchen design professional daily use*, C_8:…you *prefer*

the result for each Q_i, except for Q_5, exhibits significant differences ($p < 0.001$) in the scores in favour of GGI. Therefore, we can accept H3.

Regarding Q_5, 80% of the participants considered that the DM interface is less appropriate for one-time or occasional use. Unexpectedly, 100% of the participants, including the two professionals at the task domain, considered GGI appropriate not only for occasional use but also for frequent use. Consequently, as far as these opinions are concerned, they did not find the use of guiding obtrusive.

Results from the comparative questionnaire: the user preference

Regarding the final *questionnaire* composed of eight C_i *comparative* questions, Fig. 7 shows the percentage of participants preferring one way of interaction over the other. 100% of participants were clearly in favour of GGI. This might not be surprising among unskilled users, but it is remarkable in the case of kitchen designers, who are used to more completed and sophisticated DM interfaces.

We next quote some samples of the most repeated opinions expressed by the participants:

"… [GGI interface] liked me a lot for its reliance explaining it all… the other system [DM] leaves you sort of lost",

"It doesn't matter you don't know about it, the system [GGI] tells you all along what's next",

"I like it more the other [GGI], it's much easier; this one [DM] doesn't tell you what you have to do, you have to know it, the other guides you".

Results from a short eye tracker analysis

Aside from the main study, we conducted a quick and simple eye tracking experiment only for two participants randomly chosen. We did not try to draw important conclusions here but simply to know the most frequent screen regions the users were

looking at with each type of interface. The algorithm *k*-means was executed using the statistical package R, searching for hot regions manifested through point clustering. The optimum value of *k* for the *k*-means was obtained using the *elbow method*. This algorithm helps to automatically find, through iterative evaluation, the optimum number of clusters for a cloud of points (see, for example [48]). This analysis showed many more points and a larger number of clusters for the DM than for the GGI type of interface. The larger number of points logically corresponds with longer time periods spent whereas the double number of clusters (12 versus 6) is associated with higher complexity. This simply reinforced our study, but as no relevant conclusions could be drawn, it was not extended to more users. The experiment results have not been included in the final numerical analysis.

Discussion

As it was explained in "Materials and tasks" section, to support the study we have developed a real interface for each type of interaction. It was necessary to decide what kind of application was the most appropriate for the experiment. Firstly, the application had to be suitable for an occasional use. Secondly, it had to be appropriate for a widespread domain, valid for all ages, genders, and knowledge. And finally, it had to involve tasks requiring extensive object manipulation.

The design of a kitchen is a fairly open and well-known task which requires a lot of interactive objects for the user to manipulate. We found that all the available applications were based on conventional DM interfaces. Moreover, this is a task mostly to be performed only occasionally.

When comparing both types of interfaces, three hypotheses, H1, H2, and H3, were established. The second hypothesis (H2) refers to the number of errors. It seemed natural to expect that guiding the users would prevent them from making mistakes, needing fewer moderator assistances. The obtained results have totally proved this hypothesis: for occasional use, users who are not computer experts make fewer errors using GGI than using DM.

Our first hypothesis (H1) refers to timing: non-expert occasional users proceed quicker with GGI than with standard DM. H1 may seem counterintuitive: following a sequence of steps seems to be slower than directly manipulate elements. However, the analysis of the collected data shows that GGI is more time-efficient than DM. The only task for which there was not a significant difference was the Task 2, which being so repetitive, the user could end learning how to do it.

The last hypothesis (H3) has been first indirectly demonstrated from the answers of users in the *post-test questionnaires,* and then directly demonstrated from the final *comparative questionnaires.* These answers highlighted the fact that all the participants had a strong preference for GGI. For example, according to C_4, 100% of them consider that GGI is easier to use and requires less training than DM. The answers to the questions Q_1 to Q_4 and C_1 to C_3 showed the users felt guided on the "what to do" and the "how to do," which is one of the main goals of a GGI interface, in order to reduce the learning curve.

Regarding Q_5, being a question about the prospect of using both types of interfaces in the future, we should not take it but as an indication of satisfaction and never should it be taken as an accurate prediction. Still, it was unexpected to find that all participants, professionals included, considered GGI appropriate not only for occasional but for frequent use. Moreover, 80% of the participants considered that the DM interface is appropriate neither for one-time nor occasional use.

Domain professional users provided additional unexpected results: first, they managed to do the job sooner and with fewer errors with GGI; second, they showed a preference for that style. This preference could arise from the lower quality of our DM application compared to the commercial ones. But this must have been the case for both of our interfaces, not only for the DM one. In any case, the professional users did say that our DM interface had a close resemblance to commercial ones. This feedback from professionals referred only to the way they had worked with GGI. They did not consider necessary to allude to the quality of our DM version.

One of the limitations of GGI is its unsuitability for creative users or applications with many concurrent functionalities. GGI is not appropriate for users who interact creatively with applications, in such a way that the own interaction notably inspires and influences the final result. For example, a graphics editor could not be suitable for its use with a guided interface. Other examples could be spreadsheet, word processor or powerful video editing applications.

On the other hand, the specificity of the goals in less creative tasks is of utmost importance when it comes to guiding the interaction. Booking a room in a hotel, choosing a seat for our theatre performance, or applying specific transformations to a batch of files, are some examples where GGI is definitely appropriate. Nowadays, the advance of the Internet of Things (IoT) is making the quick use of pretty specific devices even more demanding of direct guiding [49]. The control of complex robotic tasks through touch interfaces, the use of brain-computer interfaces, especially for disabled people, does make guiding the user a necessity [50].

Finally, recent Deep Learning technologies [51] are quite promising for GGI. Adaptive user interfaces employing machine learning can create automated changes in the guiding and therefore in the complexity of the interaction.

Conclusions

The recent shifts in technology has brought about many new kinds of users who occasionally access unfamiliar systems in more and new scenarios. These are mostly users who do not want or cannot afford a learning curve. We have shown that for them, conventional *direct manipulation* is not always the most suitable interaction style. We have proposed and tested *goals-guided interaction* as a different and alternative way of interaction. Its main purpose is to strictly guide the user, step by step, in a hierarchical and progressive way, both on the "what to do" and the "how to do it," according to their current goal. This no-previous-knowledge-required approach shortens significantly the learning curve. In order to support this proposal, a quantitative and qualitative comparative study, based on user testing, was performed. The results have shown that, for non-expert occasional users, GGI is more efficient in time and less error prone than the conventional direct manipulation. Participants also showed a strong preference for

GGI. This study does not try to diminish the huge importance of DM interfaces. What it points out is that an integrated guiding system as GGI is preferred by occasional users, generally being a better option for this kind of users.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13673-020-0209-2.

Additional file 1. GGI Methodology and notation for the specification of a GGI interface.

Additional file 2. Specification of the Tested GGI interface.

Additional file 3. GGI data tables.

References

1. Shneiderman B, Plaisant C, Cohen M, Jacobs S, Elmqvist N, Diakopoulos N (2016) Designing the user interface: strategies for effective human-computer interaction. 6th edn. Pearson
2. Sharp H, Preece J, Rogers Y (2019) Interaction design: beyond human-computer interaction, 5th edn. Wiley, Indianapolis
3. Carroll JM, Thomas JC (1982) Metaphor and the cognitive representation of computing systems. IEEE Trans Syst Man Cybern 12:107–116. https://doi.org/10.1109/TSMC.1982.4308795
4. Shneiderman B (1983) Direct manipulation: a step beyond programming languages. IEEE Comput 16(8):57–69
5. Hutchins EL, Hollan JD, Norman DA (1986) Direct manipulation interfaces. In: Norman DA, Draper SW (eds) User-centered system design. Erlbaum, Hillsdale, pp 87–124
6. Moran TP (1981) An applied psychology of the user. Comput surv 13:1–11
7. Coe M (1996) Human factors for technical communicators. Wiley, New York
8. Stasko J (1996) Future research directions in human-computer interaction. ACM Comput Surv 28(4):145. https://doi.org/10.1145/242224.242410
9. van Dam A (1997) Post-WIMP user interfaces. Commun ACM 40(2):64–65
10. Norman DA (2010) Natural user interfaces are not natural interactions. Interactions 17(3):6–10
11. Selker T (2008) Touching the future. Commun ACM 51(12):14–16
12. Buxton B (2007) Multi-touch systems that I have known and loved. Microsoft Res 56:1–11
13. Nielsen J, Budiu R (2012) Mobile usability. New Riders
14. Hix D, Hartson H. Developing user interfaces: ensuring usability through product and process. New York: Wiley; 1993. 471-57813-4.
15. Nielsen J (1993) Usability engineering. Academic Press, San Diego. ISBN 0-12-518405-0
16. Clarkson J, Coleman R (eds) (2015) Special issue: Inclusive design. Applied Ergonomics, 46, part B, pp 233–324
17. Savidis A, Stephanidis C (2004) Unified user interface design: designing universally accessible interactions. Interact Comput 16(2):243–270
18. W3C-WAI. Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/. Accessed 20 Dec 2018
19. Americans with disabilities (2008) act of 1990 incorporating the changes made by the ADA Amendments Act of 2008. http://www.ada.gov/pubs/adastatute08.pdf Accessed 01 Dec 2018
20. Mueller JP (2003) Accessibility for everybody: understanding the section 508 accessibility requirements. Apress
21. Giraud S, Thérouanne P, Steiner DD (2018) Web accessibility: filtering redundant and irrelevant information improves website usability for blind users. Int J Hum Comput Stud 111:23–35
22. Charland A, Leroux B (2011) Mobile application development. Commun ACM 54(5):49–53

23. Gong J, Tarasewich P (2004) Guidelines for handheld mobile device interface design. In Proceedings of DSI 2004 Annual Meeting. pp 3751–3756
24. Sharples M (2000) The design of personal mobile technologies for lifelong learning. Comput Educ 34(3):177–193
25. Hollnagel E, Woods D (2005) Joint cognitive systems: foundations of cognitive systems engineering. CRC Press, Boca Raton
26. Huang H, Yang M, Yang C, Lv T (2019) User performance effects with graphical icons and training for elderly novice users: a case study on automatic teller machines. Appl Ergon 78:62–69. https://doi.org/10.1016/j.apergo.2019.02.006
27. Microsoft Wizards. https://docs.microsoft.com/es-es/windows/desktop/uxguide/win-wizards. Accessed 28 Jan 2019
28. Tidwell J (2011) Designing interfaces: patterns for effective interaction design. O'Reilly Media Inc., Gravenstein Highway North, p 95472
29. Williams B, Berlanstein D (2018) Creating "wizards" on the library's website. J Med Libr Assoc 106(4):490
30. Jo D, Kim GJ (2019) IoT + AR: pervasive and augmented environments for "Digi-log" shopping experience. Hum Cent Comput Inf Sci 9:1
31. Bar-Or S, Meyer J (2019) What is good help? responses to solicited and unsolicited assistance. Int J Hum Comput Interact 35(2):131–139. https://doi.org/10.1080/10447318.2018.1437866
32. Janssen CP, Donker SF, Brumby DP, Kun AL (2019) History and future of human-automation interaction. Int J Hum Comput Stud 131:99–107. https://doi.org/10.1016/j.ijhcs.2019.05.006
33. Nieto A, Rios R (2019) Cybersecurity profiles based on human-centric IoT devices. Hum Cent Comput Inf Sci 9:39. https://doi.org/10.1186/s13673-019-0200-y
34. Gillingham P (2019) Developments in electronic information systems in social welfare agencies: from simple to complex. Br J Soc Work 49(1):135–146
35. Holl K, Nass C, Vieira V, Villela K (2017) Safety–critical mobile systems–the RESCUER interaction evaluation approach. J Ubiquit Syst Pervasive Netw 9(1):1–10
36. John BE, Kieras DE (1996) The GOMS family of user interface analysis techniques: comparison and contrast. ACM Trans Comput Hum Interact 3(4):320–351
37. John BE, Kieras DE (1996) Using GOMS for user interface design and analysis: which Technique? ACM Trans Comput Hum Interact 3(4):287–319
38. Diaper D, Stanton N (2004) The handbook of task analysis for human-computer interaction. Lawrence Erlbaum Associates, Mahwah
39. Gonçalves TG, de Oliveira KM, Kolski C (2017) The use of task modeling in interactive system specification. Cogn Tech Work 19:493–515. https://doi.org/10.1007/s10111-017-0427-1
40. Kieras D (1997) A guide to GOMS model usability. evaluation using NGOMSL. In: Helander M, Landauer T (eds) The handbook of human-computer interaction, 2nd edn. North-Holland, Amsterdam, pp 733–766
41. Druryb JL, Scholtz J, Kieras D (2007) Adapting GOMS to model human-robot interaction. In: Proceedings of the ACM/IEEE international conference on Human-robot interaction (HRI '07). ACM, New York, 41–48. https://doi.org/10.1145/1228716.1228723
42. Gallagher KM, Cameron L, De Carvalho D, Boulé M (2019) Does using multiple computer monitors for office tasks affect user experience ? a systematic review. Hum Factors. https://doi.org/10.1177/0018720819889533
43. Loranger H (2016) Checklist for planning usability studies. NN/g Nielsen Norman Group. www.nngroup.com/articles/usability-test-checklist/ Accessed 25 Mar 2018
44. Nielsen J (2012) How many test users in a usability study? NN/g Nielsen Norman Group. https://www.nngroup.com/articles/how-many-test-users/ Accessed 15 Mar 2018
45. Barnum CM (2011) Usability testing essentials: ready, Set…Test! Morgan Kaufmann Publishers, Burlington, USA
46. IKEA. http://www.ikea.com/ms/en_US/rooms_ideas/kitchen_howto/NA/plan_your_kitchen_in_3d.html http://kitchenplanner.ikea.com/US/UI/Pages/VPUI.htm Accessed 06 Oct 2016
47. Rey D, Neuhäuser M (2011) Wilcoxon-Signed-Rank Test. In: Lovric M (ed) International encyclopedia of statistical science. Springer, Berlin, Heidelberg
48. Gorakala SK, Usuelli M (2015) Building a recommendation system with R. Packt Publishing Ldt. Birmingham, UK, pp 15–18
49. Bissoli A, Lavino-Junior D, Sime M, Encarnação L, Bastos-Filho T (2019) A human-machine interface based on eye tracking for controlling and monitoring a smart home using the internet of things. Sensors 19:859. https://doi.org/10.3390/s19040859
50. Kuhner D, Fiederer LDJ, Aldinger J, Burget F, Völker M, Schirrmeister RT, Do C, Boedecker J, Nebel B, Ball T, Burgard W (2019) A service assistant combining autonomous robotics, flexible goal formulation, and deep-learning-based brain–computer interfacing. Robotics Autonomous Syst 116:98–113. https://doi.org/10.1016/j.robot.2019.02.015
51. Zacharias J, Barz M, Sonntag D (2018) A survey on deep learning toolkits and libraries for intelligent user interfaces. arXiv: 1803.04818

Publisher's Note