Studying the Simultaneous Visual Representation of Microgestures

Vincent Lambert, Alix Goguey, Sylvain Malacria, Laurence Nigay

Studying the Simultaneous Visual Representation of Microgestures

VINCENT LAMBERT, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France
ALIX GOGUEY, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France
SYLVAIN MALACRIA, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, France
LAURENCE NIGAY, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France

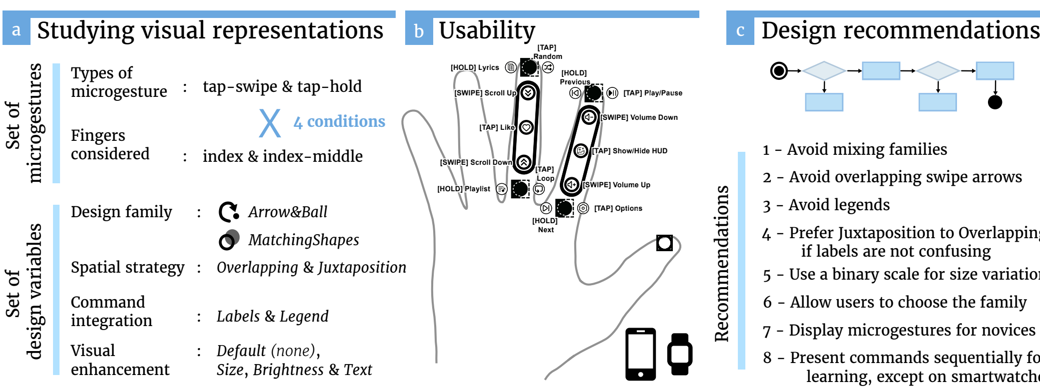


Fig. 1. Visual abstract of this paper that focuses on visual representations that simultaneously show the multiple microgestures available in an application. (a) We explore these simultaneous representations of microgestures based on 4 variables (their design, the organization of the visual cues, the presentation of the commands and additional visual enhancements) according to 4 conditions defined by the types of microgestures considered, i.e. `tap-swipe` or `tap-hold`, and the finger(s) they are performed on, i.e. `index` or `index-middle` fingers. (b) Based on the results, we design a representation for a music application, and we test its usability and suitability for wearable devices. (c) Finally, we extract 8 design recommendations from the 3 experiments.

Hand microgestures are promising for mobile interaction with wearable devices. However, they will not be adopted if practitioners cannot communicate to users the microgestures associated with the commands of their applications. This requires unambiguous representations that simultaneously show the multiple microgestures available to control an application. Using a systematic approach, we evaluate how these representations should be designed and contrast 4 conditions depending on the microgestures (tap-swipe and tap-hold) and fingers (index and index-middle) considered. Based on the results, we design a simultaneous representation of microgestures for a given set of 14 application commands. We then evaluate the usability of the representation

Authors' Contact Information: Vincent Lambert, vincent.lambert@univ-grenoble-alpes.fr, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France; Alix Goguey, alix.goguey@univ-grenoble-alpes.fr, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France; Sylvain Malacria, sylvain.malacria@inria.fr, Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, Lille, France; Laurence Nigay, laurence.nigay@univ-grenoble-alpes.fr, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France.

Vincent Lambert, Alix Goguey, Sylvain Malacria and Laurence Nigay. 2024. *Studying the Simultaneous Visual Representation of Microgestures.* Proc. ACM Hum.-Comput. Interact. 1, MHCI, Article 276 (September 2024), 34 pages.

https://doi.org/10.1145/3676523

for novice users and the suitability of the representation for small screens compared with a baseline. Finally, we formulate 8 recommendations based on the results of all the experiments. In particular, redundant graphical and textual representations of microgestures should only be displayed for novice users.

CCS Concepts: • **Human-centered computing → Usability testing**.

Additional Key Words and Phrases: Representation of microgestures; Microgesture; Command

ACM Reference Format:
Vincent Lambert, Alix Goguey, Sylvain Malacria, and Laurence Nigay. 2024. Studying the Simultaneous Visual Representation of Microgestures. *Proc. ACM Hum.-Comput. Interact.* , (September 2024), 34 pages. https://doi.org/10.1145/3676523

1 Introduction

Microgestures are swift and subtle movements that can be used to interact with an application. They define a versatile (eye-free, hand-free or not) and socially accepted interaction technique [13, 14, 36] that makes them promising for wearable computing [7, 30, 37]. Various types of microgesture exist, e.g. mouth [17] and feet [24] microgestures. However, research on thumb-to-finger microgestures is more mature and they are integrated into new industrial products like the Apple Watch [2] and the future Meta Wristband for Augmented Reality (AR) [1]. Consequently, in this paper, we consider microgestures as hand-free finger movements, independent of forearm and wrist movements [12]. As with any interactive system, all the microgestures and their corresponding commands of an application can be presented in a manual, on a website or in the application's graphical user interface. However, for effective adoption of microgesture interaction, **how should microgestures be represented?** Indeed, regardless the type of display, practitioners need information to convey all microgestures and commands to end-users in an efficient and intelligible way [33].

In any application, multiple commands (and their corresponding microgestures) are typically available at the same time [13, 45, 54, 58]. Representations of microgestures[1] can be dynamic, e.g. videos or animated visual cues [12, 23, 50], or static, e.g. sequence of drawings or expert notations [4, 12, 50]. Dynamic representations suffer limitations. First, presenting each command with animations can quickly become excessively time-consuming even with a limited number of gestures [8]. Second, users have to memorise the microgestures for future use, which might be error prone [9]. Sequential displays of static images suffer the same limitations. It makes them unsuitable for daily use with wearable technologies that require being able to quickly find and/or recall a command. Hence, we focus on techniques presenting representations of microgestures on the same visual space. Placing multiple images and descriptions side-by-side quickly reach its limits when there are many [13, 41, 59]. Another static approach consists of conveying multiple microgestures in the same graphic layout. [22, 35, 46, 52, 58]. This is the approach of SegTouch [54] and Hand-Proximate User Interfaces (HPUIs) [22, 45] which convey interactivity with fingers using widgets. This approach highlights the need to access multiple microgestures simultaneously even though they do not explicitly represent the gestures. As of now, Lambert *et al.* [35] explored design solutions to represent *a single* microgesture at a time. These design solutions provide a starting point for studying the simultaneous representation of *multiple* microgestures.

In this work, we conduct three experiments that investigate the simultaneous representation of microgestures[2]. We focus in particular on the `tap`, `swipe` and `hold` microgestures whose representations lead to a significant visual overlap when used simultaneously. `Experiment 1` evaluates

[1]In the paper, we use the expression "representation of microgesture" to refer to the visual representation of a microgesture and its associated command.

[2]In the paper, we use the expression "simultaneous representation of microgestures" to refer multiple "representations of microgestures" all arranged on the same hand image.

the trade-off between density and clarity[3] of the information (see Figure 1a). For this purpose, we consider two of the best designs proposed by Lambert *et al.* [35]: $_{ArrowAndBall}$ and $_{MatchingShapes}$. We then evaluate a simultaneous representation of microgestures with two complementary experiments. In `Experiment 2`, 14 participants use microgestures to control a music application (see Figure 1b). Finally, in `Experiment 3`, we evaluate the relative strengths and weaknesses of this representation with space-constrained devices such as smartphones and smartwatches. With the data resulting from the three experiments, we build a decision tree organized around 8 recommendations (see Figure 1c).

Our contributions are threefold:

- Strategies to design simultaneous representations of microgestures (with their corresponding commands);
- A decision tree for practitioners to inform the design of simultaneous representations of microgestures for a given application;
- Inkscape plugins to easily create representations of a single microgesture or simultaneous representations of microgestures (see Figures 1b, 2, 8, 12, 13)[4].

2 Related work

We begin by examining work on the representation of microgestures and commands. We then review the strategies for composing visual representations.

2.1 Representing microgestures

We distinguish three approaches for the static and graphical representation of microgestures.

The first approach focuses on microgestures and filters out the context of use, which results in *expert notations*. A typical example is µGlyph [11, 12], a precise hybrid (graphical and textual) notation designed to be usable after training (10 minutes). Typically, with this notation, a `tap` of the thumb on the index is written $_{t}\overset{\circ\bullet}{\blacktriangledown}(_{i}\bullet);_{t}\overset{\bullet\circ}{\blacktriangle}$ (touch; release), a `swipe` of the thumb along the index is written $_{t}\overset{\circ\bullet}{\blacktriangledown}(_{i}\bullet);(_{t}\overset{\bullet\bullet}{\blacktriangledown}(_{i}\bullet)\,|\,_{t}\overset{\bullet\bullet}{\blacktriangle}(_{i}\bullet));_{t}\overset{\bullet\circ}{\blacktriangle}$ (touch; drag; release) and a `hold` of the thumb on the index is written $_{t}\overset{\circ\bullet}{\blacktriangledown}(_{i}\bullet);_{t}\overset{\bullet}{\blacksquare}$ (touch; pause). Although highly expressive, µGlyph is aimed at experts, such as HCI researchers or UX designers, but not end-users. In addition, reading a microgesture takes 36 seconds on average [12], even after the training phase. We thus define our first design goal (**DG1**): representations of multiple microgestures should be quick and easy to use.

The second approach focuses on the context of use and employs *widget-based techniques*, intending that users will infer microgestures from the interface design. Typically, Hand Proximate User Interfaces (HPUIs) [22, 45] overlay onto users' hand various widgets that are meant to be operated via various kind of gestures. For instance, they may overlay push buttons on different phalanges or a slider on a finger. Nevertheless, this technique is not suitable for displaying several commands on the same finger zone as it requires widgets that visually communicate several user inputs, which is a significant challenge for GUI design [39, 40]. In addition, HPUIs have only been tested in an Augmented or Virtual Reality context and rely a lot on user's interpretation. It is unclear how they convey that users cannot and should not interact with these widgets using their other hand [29]. This leads to our second design goal (**DG2**): representations of multiple microgestures should make it possible to represent microgestures sharing the same interacting zones simultaneously.

Finally, the third approach considers both the microgestures and their context of use, which results in *interaction illustrations* [3]. Various examples exist in the literature. Lambert *et al.* [35]

[3]As used by Quispel *et al.* [47], Chen [15] and Haapio [42] for visualization.
[4]The plugins and representations used in the experiments are available as supplementary material.

	DG1 quick and easy to use	DG2 allows shared interacting zones	DG3 easily scales with the number of microgestures
µGlyph **Chaffangeon Caillet *et al.* [12]**	✗	✓	✓
HPUIs **Faleel *et al.* [22]**	✓	✗	✓
Single-picture representations **Lambert *et al.* [35]**	✓	✓[5]	✗
Simultaneous representations	✓	✓	✓

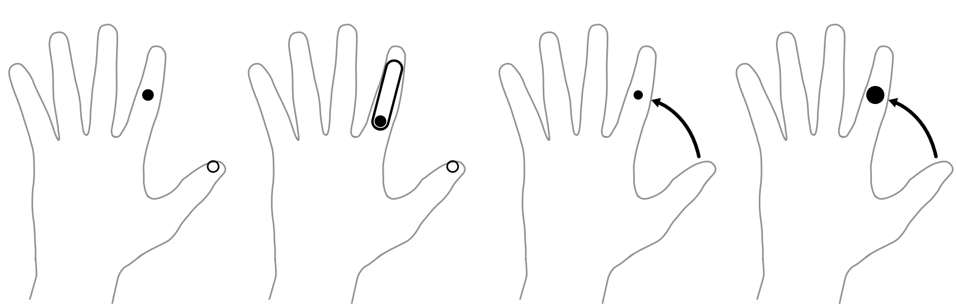
Table 1. Comparison of existing visualization strategies with their main strengths and weaknesses.

considered them and represented 4 microgestures, namely `tap`, `swipe`, `hold` and `flex`, using groups of consistent designs called "families". Families differ by showing or not trajectory, direction, actuator and receiver fingers. However, they were evaluated for a single microgesture at a time and without associated commands. Displaying various single-picture representations on the same layout would quickly become unreadable with small screen sizes. This leads to our third design goal (**DG3**): representations of multiple microgestures should easily scale while remaining practical.

Table 1 summarizes the strengths and weaknesses of current visualization techniques in relation to the simultaneous representation of microgestures presented in the following pages.

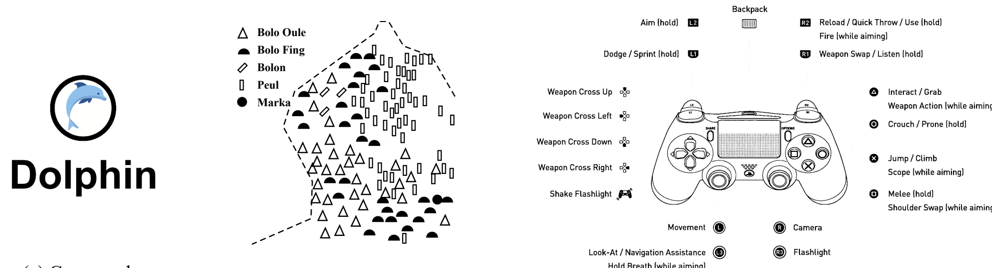
By considering the most common microgestures, `tap` and `swipe` representations can *partially* use the same finger zone depending on the family (see Figure 2a and 2b). It would cause overlapping that we anticipate as detrimental for users' understanding. For `tap` and `hold` representations, the problem is even more important as they use the same finger zone and thus *completely* overlap, e.g. the $_{ArrowAndBall}$ family differentiates a `tap` from a `hold` using a small and large ball respectively (see Figure 2c and 2d). Lambert *et al.* [35] also reported that the family showing transparent fingers caused visual overload which makes it unsuitable for the simultaneous representation of

[5]With single-picture representations, depicting microgestures sharing the same interacting zone requires the use of several hand drawings with spatial or temporal multiplexing.



(a) $_{MaS}$ tap (b) $_{MaS}$ swipe (c) $_{A\&B}$ tap (d) $_{A\&B}$ hold

Fig. 2. (a) `tap` and (b) `swipe` representations for the $_{MatchingShapes}$ family. (c) `tap` and (d) `hold` representations for the $_{ArrowAndBall}$ family.



(a) Command icon and label

(b) Map with its legend

(c) In-game control layout for *The Last of Us™ - Part II*

Fig. 3. Semantic link techniques: (a) label close to the command icon, (b) legend integrated in the original visualization with repeated glyphs (figure inspired from Bertin's work [6]), (c) additional label distinguishing two commands on the same button

multiple microgestures. These are crucial aspects of visualization for the clarity of microgesture representations. We thus choose to consider two sets of microgestures, i.e. **taps with swipes** and **taps with holds** and **two families**, i.e. $_{ArrowAndBall}$ which explicitly shows the trajectory for each microgesture and $_{MatchingShapes}$ which does not show any trajectory.

2.2 Presenting commands in existing applications

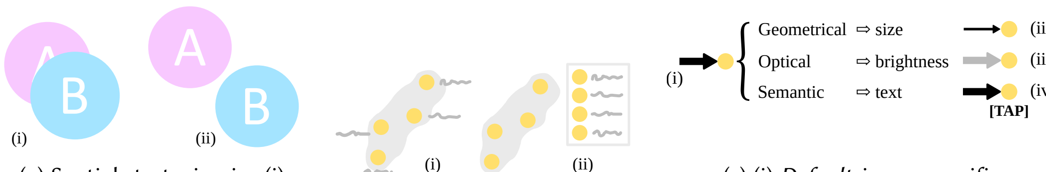
In line with best practice, commands are presented in the form of icons with text [10] (see Figure 3a). Yet, the same zone, e.g. a physical button, can be related to multiple commands which could provoke confusion. This issue is common to diverse fields and is often handled with crib-sheets [4, 56] or legends (see Figure 3b). In video-games, control panels (see Figure 3c) are more frequent than crib-sheets. They take advantage of the controller itself to display multiple commands simultaneously. They often use an icon or a textual information to distinguish 2 commands triggered by the same button, e.g. in Figure 3c the L1 button can be pressed to dodge or held to sprint. In the literature, we did not find any guidelines explaining how to display multiple commands that can be associated with the same button or physical zone.

With the hand as a controller, different microgestures can involve the same physical zone, i.e. a swipe use multiple finger joints which can be tapped or held. Their representations thus may interfere with each other when displayed simultaneously. Therefore, the following section focuses on visual composition strategies to enhance the clarity of a simultaneous representation of microgestures.

2.3 Spatially composing visual representations

Several studies focus on the visual composition of visual objects, e.g. two graphs. Four strategies appear to be predominant: *Overloading*, *Nesting*, *Superimposition* and *Juxtaposition* [25, 31, 38, 53]. *Overloading*, *Nesting* and *Superimposition* concern cases where two visual objects overlap to varying degrees. *Juxtaposition*is different as it requires to separate the visual objects. Positioning all visual elements side by side, without overlap, takes up more space, but ensures that each of them cannot be misinterpreted. Finally, a fifth strategy is particularly predominant with graphs: explicit encoding [25, 38]. It uses computation to determine and depict a relationship between data points, e.g. creating a new variable $x - y$ or $x \times y$ encompassing both x and y.

Multiple visual objects involve visual cues that are not self-explanatory, e.g. a map can use special glyphs to depict town halls. In that case, each visual cue needs to be explicitly linked to a



(a) Spatial strategies, i.e. (i) Overlapping and (ii) Juxtaposition, for microgesture representation.

(b) Preferable use of (i) *Labels* or a common (ii) *Legend* for command presentation.

(c) (i) *Default*, i.e. no specific enhancement, (ii) *Size* or (iii) *Brightness* modification and (iv) *Textual* annotation depending on the microgesture type.

Fig. 4. Spatial strategies relevant to (a) microgestures or (b) the command presentation and (c) considered visual enhancements.

semantic or iconic definition to be understood. This ties in with the *Integration* proposed by Javed and Elmqvist [31] whose idea is to add explicit linking of distinct data items in the specific case of juxtaposed visualizations. Practitioners already integrate this concept in their visualizations. With schemes, they usually use close labels or lines (see Figure 3a), whereas with maps [26] or glyph-based graphs [21] legends are ubiquitous (see Figure 3b).

Practitioners also take advantage of additional enhancements to implicitly mark the difference between two categories of visual cues. These enhancements are derived from the visual channels that define each visual cue. Chen *et al.* [28] proposed 4 classes of visual channel based on Bertin's work [6]: Geometric, Optical, Relational and Semantic. These classes gather more than 30 channels such as size (Geometric), brightness (Optical), distance (Relational) and text (Semantic). With these classes and the visual strategies introduced earlier, we can now define the relevant design strategies for simultaneous representations of microgestures.

3 Design strategies to simultaneously represent microgestures

3.1 Selecting spatial strategies

Representing a microgesture and presenting the associated command are two orthogonal problems.

The representation of microgestures is related to the 4 strategies for visual composition defined by Javed and Elmqvist [31]. We choose to consider together the strategies that use different degrees of overlap, namely overloading, nesting and superimposition. We therefore group them under the term *Overlapping* (Figure 4a-i). Concerning *Juxtaposition* (Figure 4a-ii), shifting visual cues could confuse users on where the microgesture must be performed. To build consistent representations with satisfying clarity, we need a general rule on how to juxtapose. We choose to shift visual cues having different meanings, e.g. a disk showing the receiver of a `tap` and a stroke showing the trajectory of a `swipe`, but not visual cues that convey the same meaning, e.g. both trajectory lines of two opposite arrows.

The presentation of the commands relies on the *Integration* [31] strategy previously described. We choose to use either close *Labels* or a *Legend* (see Figure 4b).

These spatial strategies are promising to represent multiple microgestures simultaneously in both *partially overlapped*, i.e. `taps` with `swipes`, and *completely overlapped* contexts, i.e. `taps` with `holds`. We also consider additional visual enhancements that shape a design space orthogonal to that defined by spatial strategies.

3.2 Selecting visual enhancements

Visual variables fall into 4 classes [28]: Geometric, Optical, Semantic and Relational. Relational variables (such as distance and closure) interfere with the above spatial strategies. We therefore consider only the first three classes of visual variables. As shown in Figure 4c, we consider three variables, each belonging to a distinct visual channel: *Size* (Geometric), *Brightness* (Optical) and *Text* (Semantic). *Size* was already identified by Bertin as a major visual variable [6] but he also mentioned value, i.e. *Brightness*, to be one of them too. We choose to focus on *Brightness* instead of color to make the study inclusive towards color-blind people. If *Size* and *Brightness* directly modify the visual cues of a microgesture representation, *Text* does not. It consists in adding a semantic information (here using brackets, see Figure 4c-iv) to the already shown command name. Together, they form a single label containing both the microgesture to be performed and the associated command.

3.3 Resulting designs

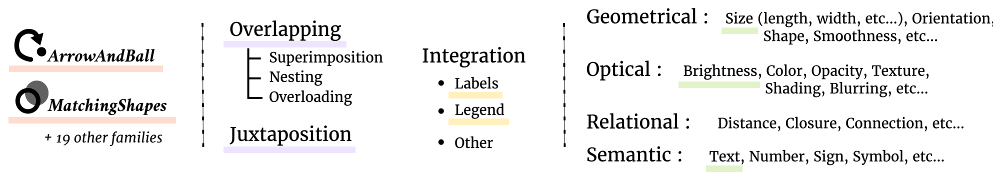
Figures 5, 6 and 7 summarize the spatial strategies and visual enhancements we have adopted to design the simultaneous representations of microgestures and their associated commands[6] used in `Experiment 1`. Enhancements aim to reinforce the distinction between two types of microgestures. As *Size* enhancement, we choose to decrease the size of the visual cues representing the `swipe` microgesture. As *Brightness* enhancement, we choose to use grey for the visual cues representing the `swipe` and `hold` microgestures. Finally, the difference between `tap` and `hold` microgestures relies on geometric channels for many families [35]. The *Size* enhancement thus interferes with the perception of `tap` and `hold` representations. Consequently, we do not consider the *Size* enhancement for `tap-hold` representations which is translated by a crossed-out cell in Figure 7.

To produce the representations schemed by Figure 6, we built a set of 5 Inkscape plugins. They have been designed to be usable by other practitioners and are available as supplementary material. Every simultaneous representation of microgestures in this paper have been computed with these tools (see Figures 1c, 2, 8, 12, 13). We choose to highlight the *Text* enhancement, i.e. the microgesture type, with a pink color in Figure 6 and Figure 7 to make the difference clear with the command name. However, please note that in real representations, the *Text* enhancement and the command name share the same color.

3.4 Study plan

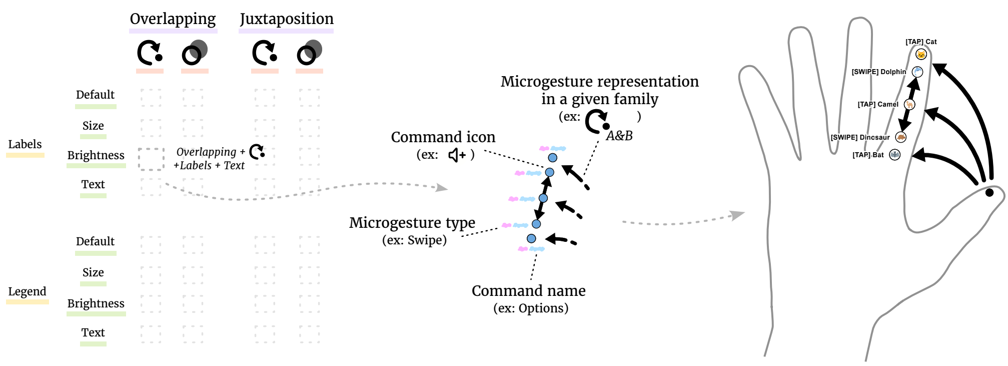
First, we need to determine *how* to design a simultaneous representation of microgestures. As a result, we choose to conduct a first experiment (`Experiment 1`), centered on how to design the best possible simultaneous representations of microgestures. With the results, we are then able to design a help display that could be relevant to a real application. However, practitioners will not use simultaneous representations of microgestures if they are not usable and are not a relevant alternative to existing representations. To evaluate these aspects, we conduct two complementary experiments. `Experiment 2` tests if the representation is actually usable in-context even by novice users. `Experiment 3` evaluates if this type of representation is indeed more adapted to wearable devices than a baseline.

[6]The default superimposition for complete overlap has been designed in light of type J multiplexing as defined by Chen *et al.* [16].



(a) Design families (b) Spatial composition (c) Command presentation (d) Visual enhancements

Fig. 5. Design strategies in relation to the main related works : (a) Lambert *et al.*[35], (b-c) Javed and Elmqvist [31], (d) Chen *et al.* [28]. The underlined ones are those considered in this study.



(a) Design space of the strategies used to represent multiple microgestures (b) Symbolic description of a representation (c) Simultaneous representation of microgestures

Fig. 6. Organization of the design space outlined in Figure 5. Stroke colors refer to Figure 5. Each strategy of the (a) design space can be described by a (b) symbolic description (see Figure 7). These are associated to (c) simultaneous representation of microgestures.

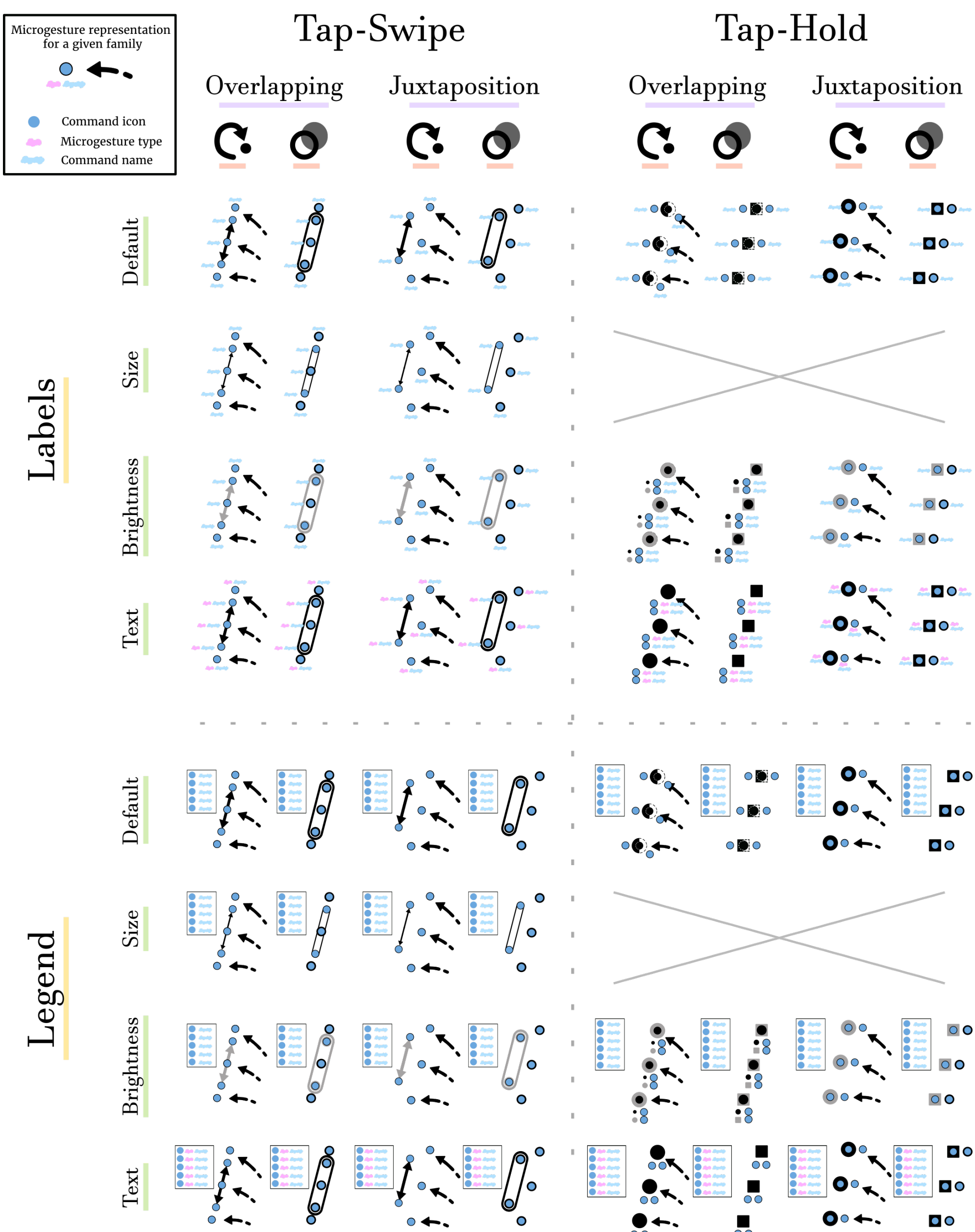


Fig. 7. Representing the `tap-swipe` and `tap-hold` conditions with the 2 families, i.e. $_{ArrowAndBall}$ and $_{MatchingShapes}$, the 2 spatial strategies, i.e. *Overlapping*, *Juxtaposition*, the 2 integration techniques, i.e. *Labels*, *Legend*, and the 4 visual enhancements, i.e. *Default*, *Size*, *Brightness* and *Text*. Stroke colors refer to Figure 5 and Figure 6.

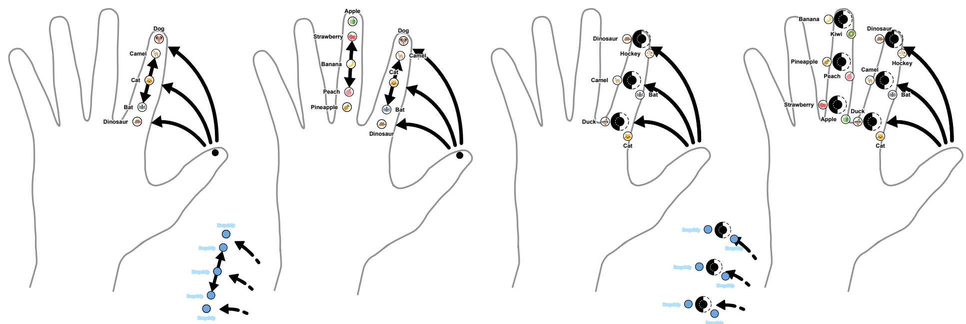


Fig. 8. Examples of $_{ArrowAndBall}$ representations with the *Overlapping* strategy with *Labels* and no enhancement for the following conditions : (a) `tap-swipe` with `index` (TS_I), (b) `tap-swipe` with `index-middle` (TS_IM), (c) `tap-hold` with `index` (TH_I) and (d) `tap-hold` with `index-middle` (TH_IM).

4 Experiment 1: Comparing design strategies

In this work, we aim to evaluate design strategies to visually represent multiple microgestures simultaneously while maximizing clarity. According to existing elicitation [13, 36, 41] and interaction studies [45, 54, 57, 58], microgesture-based applications could make use of different fingers or finger zones to propose different commands. The two most popular are the `index` finger and the `middle` finger involved in 67% of the microgestures of Chan *et al.*'s consensus set [13]. Thumb-to-finger interactions with the `index` finger and the `middle` finger have been labeled as "very easy to reach" by Dewitz [19]. Therefore, and in order to conduct an experiment of reasonable duration, we study thumb-to-finger microgestures with the `index` only and with both the `index` and `middle` fingers. In line with existing studies, we focus on 3 zones for the `tap` and `hold` microgestures [12, 19, 22, 30, 45] and 2 vertical directions for the `swipe` microgesture [12, 22, 45, 52] (see Figure 8).

Figure 9a lists the independent variables considered in this experiment. Between-subject variables concern the microgestures considered (`tap` and `swipe` or `tap` and `hold`) and the finger used (`index` finger only or `index` and `middle` fingers). Each of the four conditions was assigned to a different group of participants.

4.1 Participants & Apparatus

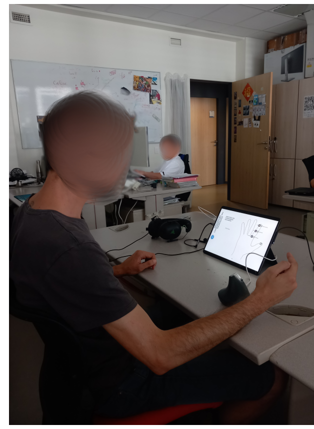
We recruited 32 volunteers (8 randomly assigned to each between-subject condition) ranging in age from 19 to 61 years old (M=28, σ=10). 16 self-identified as women, 15 as men, 1 as non-binary. They covered a wide range of professions, from physiotherapists to teachers and design students. They were recruited via social networks and on the local campus. Each participant received a 15€ gift card for taking part in the study.

The experiment setup was composed of a Microsoft Surface 9 running the experiment website (see Figure 9b). Participant used their hand to perform the microgestures and had access to the mouse to fill the form. The experimenter used a remote control on his smartphone to enter the microgestures performed by the participants (see Figure 9c).

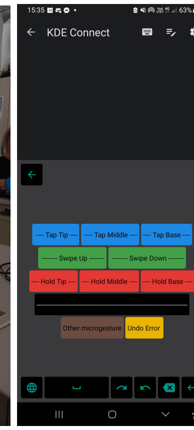
		Microgesture combo*	
		tap-swipe	tap-hold
Finger combo*	index	5µgs	6µgs
	index-middle	10µgs	12µgs
Spatial strategy		2 cases	
Integration		2 cases	
Enhancement		4 cases	3 cases
Family		2 cases	
Total		**32reps/p**	**24reps/p**

		Microgesture combo*	
		tap-swipe	tap-hold
Finger combo*	index	160µgs/p	144µgs/p
	index-middle	320µgs/p	288µgs/p

(a) Design of the experiment. "µgs" stands for microgestures, "reps" for representations and "p" for participant.
* represents a between-subject variable. Each combination of these variables corresponds to a group of 8 participants.



(b) Photo of the experiment setup.



(c) Experimenter's remote control.

Fig. 9. (a) Design of the experiment and (b-c) captures of the setup

4.2 Procedure

The experiment was based on the following procedure, which took about 1 hour to complete:

(1) Consent form and preliminary questions
(2) Video introducing the experiment and the 6 representation designs used the experiment, i.e $_{ArrowAndBall}$ `tap`, `swipe`, `hold` and $_{MatchingShapes}$ `tap`, `swipe`, `hold` representations.
(3) Training phase
(4) Evaluation phase where participants evaluate the strategies corresponding to their assigned condition (`tap-swipe` or `tap-hold` and `index` or `index-middle`)
  (a) Sequential execution of the requested commands[7] and evaluation of their difficulty[8]
  (b) Evaluation of the representation with a 5-point Likert scale[9] on 3 topics: visual overload, appeal and speed of understanding, i.e. readability.

During the training phase, the participants were randomly exposed, for both families, to all 8 single representations of the microgestures considered in the experiment, i.e. `tap` (tip, middle and base), `swipe` (up and down), `hold` (tip, middle and base). They had to correctly execute the resulting 16 microgestures twice in a row to ensure that all the participants shared the same level of expertise on the single-picture representation of each microgesture.

As shown by Figure 9a, for the `tap-swipe` conditions, participants were asked to evaluate 32 representations (2 spatial strategies × 2 integration cases × 4 visual enhancements × 2 families). With each representation, they had to execute 5 or 10 commands, i.e. 32 × 5 = 160 or 32 × 10 = 320 in total, depending on the `index` (TS_I, see Figure 8a) or `index-middle` (TS_IM, see Figure 8b) conditions respectively. For the tap-hold conditions, they were asked to evaluate 24 representations

[7]We have chosen to pick the set of commands used by Grossman *et al.* [27] with modernized icons.
[8]We opted for a binary evaluation to take into account the instantaneous feelings of the participants, i.e. easy or hard.
[9]To perform our analysis, the Likert scale from *Strongly disagree* to *Strongly agree* was encoded using multiples of 0.25 between 0 and 1 inclusive.

(2 spatial strategies × 2 integration cases × 3 enhancement strategies × 2 families). With each representation, they had to execute 6 or 12 commands, i.e. 24 × 6 = 144 or 24 × 12 = 288 in total, depending on the `index` (TH_I, see Figure 8c) or `index-middle` (TH_IM, see Figure 8d) conditions respectively. Each representation is a different combination of 4 independent variables, all having the same importance in our experimental design. Consequently, it was infeasible to counter-balance the conditions without dramatically increasing the number of participants needed. As a result, we chose to mitigate the potential sequencing effect by randomising the order of presentation.

4.3 Results

Following the recommendations for "Fair statistical communication in HCI" [20], we use 95% confident intervals computed using bootstrapping, with N the size of the considered sample. Please note that in the following, we only aggregate between-subject variables, i.e. TS vs. TH and I vs. IM, for which statistical tests of dependence reveal no significant effect. We also made sure that the absence of samples for the size enhancement in the TH condition did not impact these aggregations. Our samples are too small to apply a law of large numbers and then use parametric methods. Following the recommendations of J.O. Wobbrock and M. Kay [48], we thus use non-parametric methods, i.e. Mann-Whitney (MW) with Common Language Effect Size (CLES) [44, 55] as effect size and Aligned Rank Transform (ART) with n_p^2 as effect size.

4.3.1 Interest in the proposed user interface design.

QUANTITATIVE DATA: We observed that the participants tend to have a higher appeal for the technique in the `tap-swipe` condition than in the `tap-hold` condition (Median (Mdn)=0.7, CI=[0.65, 0.78] for TS with N=16; Mdn=0.57, CI=[0.48, 0.76] for TH (N=16); MW with p=0.0647, CLES=69%)[10].

INTERVIEW DATA: Once the training phase has passed, most participants' comments were neutral towards the technique itself ("It looks like furniture assembly instructions" [P14], "everyone could define their own [design] preference" [P32]). Nonetheless, we also received positive and insightful comments ("Having all the gestures at the same time is quite cool" [P18], "it's quite pleasant" [P22] and "A delete button in less accessible areas would be useful for critical options" [P26]).

[10] For further details, see Appendix B, Table 3b.

		Microgestures combo	Tap-Swipe				Tap-Hold			
		Spatial strategy	Overlapping		Juxtaposition		Overlapping		Juxtaposition	
		Family	A&B	MaS	A&B	MaS	A&B	MaS	A&B	MaS
Finger combo	Enhan-cement	Integration								
Index	Default	Labels	0.72	0.82	0.75	0.65	0.83	0.92	0.98	0.92
		Legend	0.72	0.7	0.68	0.75	0.81	0.88	0.9	0.79
	Size	Labels	0.8	0.6	0.78	0.88				
		Legend	0.72	0.78	0.72	0.78				
	Brightness	Labels	0.72	0.8	0.7	0.72	0.92	0.9	0.92	0.88
		Legend	0.7	0.75	0.72	0.72	0.88	0.77	0.92	0.92
	Text	Labels	0.75	0.78	0.72	0.88	0.94	0.85	1.0	0.96
		Legend	0.72	0.72	0.7	0.88	0.96	0.96	0.92	0.96
Index-Middle	Default	Labels	0.74	0.86	0.74	0.82	0.82	0.98	0.99	0.98
		Legend	0.71	0.76	0.69	0.84	0.96	0.95	0.95	0.99
	Size	Labels	0.7	0.66	0.68	0.76				
		Legend	0.75	0.79	0.68	0.76				
	Brightness	Labels	0.78	0.71	0.69	0.69	0.99	0.99	0.97	0.86
		Legend	0.6	0.79	0.71	0.71	0.97	0.96	0.93	0.96
	Text	Labels	0.71	0.89	0.69	0.82	0.96	0.98	0.97	0.98
		Legend	0.74	0.88	0.7	0.9	0.96	0.94	0.99	0.96

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

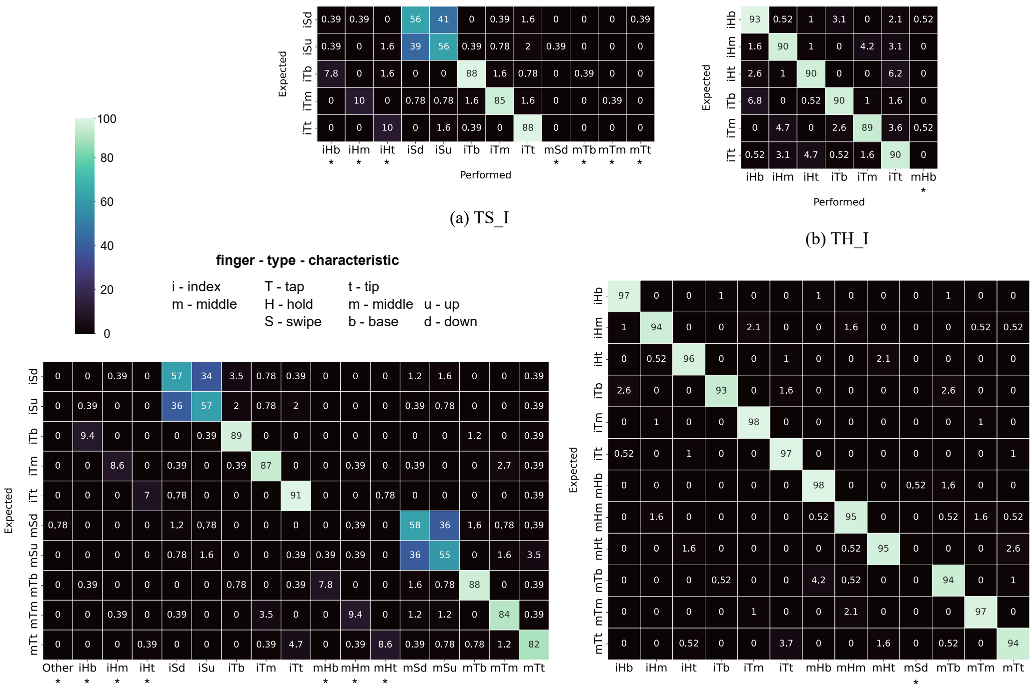
(a) **Success**

		Microgestures combo	Tap-Swipe				Tap-Hold			
		Spatial strategy	Overlapping		Juxtaposition		Overlapping		Juxtaposition	
		Family	A&B	MaS	A&B	MaS	A&B	MaS	A&B	MaS
Finger combo	Enhan-cement	Integration								
Index	Default	Labels	0.28	0.34	0.25	0.12	0.47	0.38	0.19	0.12
		Legend	0.22	0.28	0.19	0.19	0.53	0.53	0.47	0.31
	Size	Labels	0.34	0.16	0.16	0.06				
		Legend	0.19	0.16	0.25	0.19				
	Brightness	Labels	0.25	0.12	0.12	0.03	0.53	0.53	0.09	0.28
		Legend	0.22	0.19	0.22	0.22	0.62	0.72	0.44	0.47
	Text	Labels	0.47	0.5	0.47	0.25	0.56	0.38	0.31	0.22
		Legend	0.5	0.44	0.44	0.31	0.62	0.47	0.53	0.53
Index-Middle	Default	Labels	0.31	0.28	0.28	0.28	0.41	0.31	0.34	0.22
		Legend	0.56	0.59	0.62	0.47	0.56	0.62	0.47	0.41
	Size	Labels	0.28	0.19	0.38	0.16				
		Legend	0.56	0.44	0.66	0.53				
	Brightness	Labels	0.34	0.09	0.31	0.31	0.25	0.41	0.25	0.22
		Legend	0.56	0.53	0.66	0.5	0.75	0.72	0.59	0.32
	Text	Labels	0.66	0.5	0.78	0.59	0.62	0.62	0.81	0.56
		Legend	0.62	0.56	0.75	0.53	0.72	0.66	0.62	0.53

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

(b) **Overload**

Table 2. Mean value of the success and overload measures for each representation among the 8 participants that have seen it



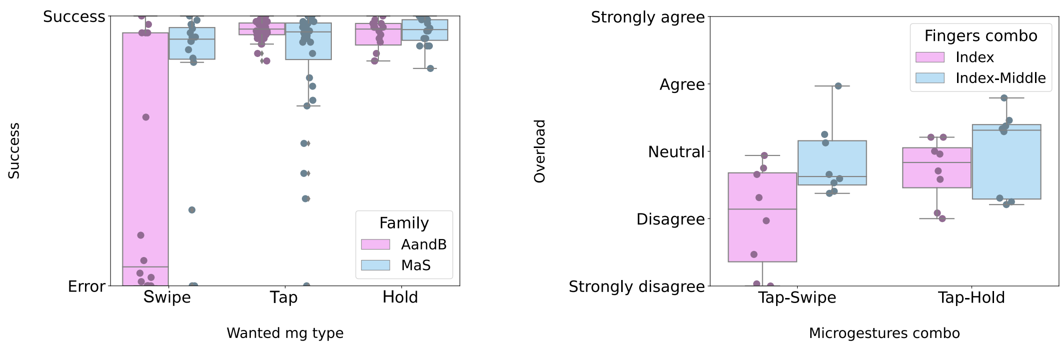
(a) TS_I (b) TH_I (c) TS_IM (d) TH_IM

Fig. 10. Performed microgesture depending on the expected microgesture (per condition). Each microgesture is labelled is labelled with 3 letters according to the "finger-type-characteristic" structure. Examples: **iTt** stands for **index tap tip** and **mSd** for **middle swipe down**. The * sign highlights the unexpected microgestures that have been performed in each condition.

4.3.2 *Task success.*

Quantitative data: Overall, all the participants correctly executed the expected microgestures (Mdn=0.92, CI=[0.81, 0.97] for TS_I (N=8); Mdn=0.92, CI=[0.81, 0.97] for TH_I (N=8); Mdn=0.73, CI=[0.59, 0.9] for TS_IM (N=8); Mdn=0.96, CI=[0.94, 0.97] for TH_IM (N=8)) (see Table 2a) without errors that might have resulted from a confusion between icons (see Appendix A). However, we observe that errors are more frequent in the `tap-swipe` condition than the `tap-hold` condition (Mdn=0.75, CI=[0.7, 0.8] for TS (N=16); Mdn=0.95, CI=[0.92, 0.97] for TH (N=16); MW with p=0.00020, CLES=89%). Figure 10 and Figure 11a show that these errors are mostly related to the direction of the `swipe`. By distinguishing families (see Figure 11a), we observe that these errors are mainly found for the $_{ArrowAndBall}$ family. The very large CI for the median (Mdn=0.07, CI=[0, 0.94] for TS (N=16)) suggests that the tested $_{A\&B}$ representation of the `swipe` microgesture becomes ambiguous for indicating the direction when arrows pointing in opposite directions overlap. A smaller proportion of errors came from `taps` misunderstood for `holds`. Figure 11a shows that for the `tap` microgesture, it is the design of the $_{MatchingShapes}$ family that seems to confuse some participants.

Interview data: During the training, only one participant had issues with the $_{A\&B}$ `swipe` representation compared to 24 for the $_{MaS}$ `swipe` representation. They felt that the $_{MaS}$ `swipe` representation did not convey direction explicitly enough. Yet, almost all the participants orally expressed a confusion about $_{A\&B}$ `swipes` during the experiment ("it's not intuitive at all, we need



(a) Success for the expected microgesture type per family.

(b) Overload of the representation per microgesture combo and finger combo respectively.

Fig. 11. Reported scores of the representations with their respective quartiles.

to choose a direction" [P21], "it's unclear about what we have to do" [P24], "it's symmetric, therefore hard to compare" [P29]) whereas the $_{MaS}$ family has been strongly preferred by 8 participants out of the 16 of the `tap-swipe` conditions ("like on iPhone" [P14, P15], "more intuitive" [P14, P15, P21]). Their confusion stemmed from whether the `swipe` movement should start *from* the icon or be directed *towards* it. Out of the 16 participants of the `tap-swipe` conditions, 9 swiped *from* the icon for every `swipe` representation, 3 swiped *towards* it and 4 did both depending on the family. Even though they did both, P18 commented that it is better "to have the icon where the finger should touch" because "you don't have to think about `tap` or `swipe`, you answer the questions while doing the gesture" [P18]. This is in line with the 9 participants out of 16 who indicated they had wanted to touch then `swipe` the icon meant to represent the middle `tap` for the $_{A\&B}$ family in *Overlapping* representations. Regarding the confusion on $_{MaS}$ `tap` representation, the participants indicated it was caused by the icon border width ("tendency to mix [microgestures]" [P19], "I only look at the circle and its weight" [P22]). Indeed, the `tap` representation of the $_{MaS}$ family has a slightly bigger border than the one of the $_{A\&B}$ family and it has been confused with the `hold` representation of the $_{A\&B}$ family which has a very thick border.

4.3.3 *Causes of visual overload.*

Quantitative data (focus on between-subject variables): As expected, the overload of a representation is proportional to the number of microgestures represented which depends on the microgestures and the fingers considered (ART for TS and TH with p=0.073, n_p^2=0.058≃0.06, medium effect size ; ART for I and IM with p=0.032, n_p^2=0.084>0.06, medium effect size) (see Figure 11b).

Interview data (focus on between-subject variables): The participants confirmed this analysis ("half as much information would not cause visual overload" [P14], "overloaded because there are 24 pieces of information with redundancy" [P8]). In the `index-middle` conditions, 1 in 3 participants pointed out that this feeling of overload was intensified by the fact that the labels sometimes overlapped the wrong finger ("[I prefer] nothing in the inter-digital space" [P14], text on another finger is "disturbing" [P17]).

Quantitative data (focus on independent variables): We also observe that using *Overlapping*, a *Legend* or a *Text* enhancement causes more overload than using *Juxtaposition*, *Labels* or other enhancements respectively[11] (ART for spatial strategy with p=0.046, n_p^2=0.036>0.01, small effect size ; ART for integration with p=0.0053, n_p^2=0.07>0.06, medium effect size ; ART for enhancements with p=7.8e-05, n_p^2=0.14, large effect size).

Interview data (focus on independent variables): The participants were almost unanimous in explaining that *Legend* integration and *Text* enhancement were direct causes of overload. 10 and 9 participants out of 32 reported they have ignored the *Legend* integration and the *Text* enhancement respectively for most of the experiment. 12 participants out of 32 told that the *Legend* slows down the interaction because it causes unnecessary "round trips" [P27, P28] ("it's much better if the text is directly next to it" [P21], "[*Labels* are] so much easier" [P29], "it's useless" [P9]). 13 participants out of 32 described the *Text* enhancement as "redundant" [P15, P29] and consequently "useless" [P8, P31, P32] considering that the microgestures are already indicated by their corresponding visual cues ("there is no possible mistake" [P27], "the design code must be self-sufficient" [P31], "if there are 2 dimensions, one is superfluous" [P29]). However, 13 participants out of 32 also found that the *Text* enhancement could be a good help as it is naturally "less ambiguous" [P5]. P22 explained it would be "easier for a first time" and P15 added it "could be used as a cheatsheet, but after the training, everything is clear".

4.3.4 Evaluating spatial strategies.

Quantitative data: We previously mentioned that *Juxtaposition* caused slightly less overload than *Overlapping*. For the `tap-hold` conditions, *Juxtaposition* has also been considered more appealing (ART for appeal with p=0.067, n_p^2=0.061>0.6, medium effect size), readable (ART for readability with p=0.00035, n_p^2=0.23>0.14, large effect size) and easy to understand (ART for easiness with p=0.00018, n_p^2=0.25>0.14, large effect size). Tables to validate these assertions are available in Appendix B, and details of the calculations are available in the supplementary material.

Interview data: 11 participants out of 32 strongly expressed a preference for *Juxtaposition* against only 3 for *Overlapping*. The remaining participants had a nuanced opinion depending on the other variables. Those who preferred *Juxtaposition* explained that it was "easier" [P4] and allowed to "better understand" [P4, P11] the microgestures. Despite the training, 9 participants out of the 16 of the `tap-swipe` conditions told us that, at some point, they wanted to "swipe" the visual cue meant to represent the `tap` on the middle phalanx ("it's disturbing to have icons in the middle" [P15]), especially with the $_{A\&B}$ family. We also noticed that with the *Juxtaposition* strategy, 13 participants out of 32 systematically adapted the location of the performed microgestures to the positions of shifted visual cues, e.g. they held the *inside*[12] of the `index` finger and tapped on the *outside*[13] of the `index` finger.

4.3.5 Focus on families.

Quantitative data: Besides the issue with $_{A\&B}$ `swipe` direction and $_{MaS}$ `tap` confusion, we did not observe any significant difference between families for any of the measures.

Interview data: Nonetheless, 9 participants out of 32 and 12 participants out of 32 strongly expressed a preference for $_{A\&B}$ and $_{MaS}$ respectively. Those who preferred $_{A\&B}$ found $_{MaS}$ `swipes` "less readable" [P2] and "less clear" [P30] as they felt more confident with arrows. P17 and P30 quickly forgot the training and felt lost with having both round and square shapes to distinguish `tap` and `hold` microgestures. P9 also described $_{MaS}$ `holds` as difficult to interpret due to the "need to think". In comparison, $_{A\&B}$ was perceived as having "not any ambiguity" [P31] and being "very

[11] The all-in-one software provided as supplementary material comes with a Preset.csv file, enabling the reader to obtain more detailed information on these statements.
[12] *ulnar* side as defined by Soliman [52]
[13] *vulnar* side as defined by Soliman [52]

clear" [P1]. On the contrary, participants who preferred $_{MaS}$ clearly expressed their preference for having both round and square shapes to distinguish `tap` and `hold` microgestures in comparison with the variation in circle size used by the family $_{A\&B}$ ("squares are better than circles" [P6], "[it's] more harmonious" [P12]). They also appreciated $_{MaS}$ `swipe` as "you know where you start and where you end" [P11]. The 11 remaining participants were nuanced between the representations, sharing the opinions of both groups. Finally, 10 participants out of 32 distributed across the three groups told us that they found that $_{A\&B}$ "[tap] arrows cause more interference than they help" [P22] and were "redundant thus useless" [P11]. Similarly, P29 described as useless the disk on the thumb which explicitly shows the `swipe` actuator for the $_{A\&B}$ family ("I always use my thumb, i don't need the point or the arrows" [P29]).

4.3.6 Correlation between measures. We observe a correlation between the measures. Overload is inversely proportional to both readability (Pearson with p=1.4e-08, r=-0.49, CI=[-0.61, -0.34]) and appeal (Pearson with p=1.8e-13, r=-0.61, CI=[-0.71, -0.49]) whereas appeal is proportional to readability (Pearson with p=1.0e-07, r=0.46>0.3, CI=[0.31, 0.59]) (see Appendix B). During the interview, only P23 made an explicit link between overload and appeal.

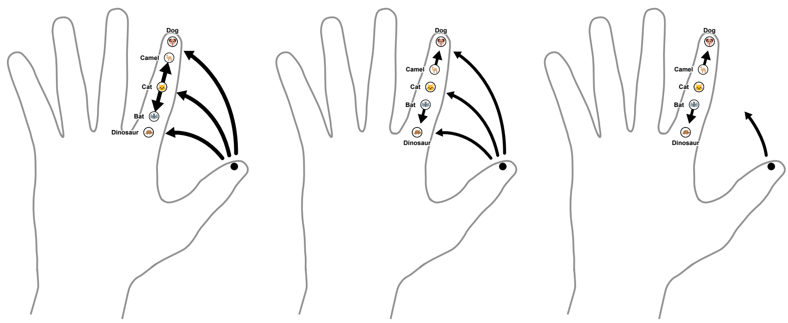
4.3.7 Participant-dependant observations. During the experiment, 4 participants reported physical pain due to repetitions, especially with the `tap` on the base of the fingers ("tendonitis if you do it all your life" [P10]). The experiment has also been described as a "dyslexia and dyspraxia exercise" by P10 and P27. On another topic, P26 verbalized they would first determine the "finger", i.e. the receiver, then the "zone", i.e. the characteristic, and then the "action", i.e. the microgesture. However, even if the participants understood the microgesture they had to do, sometimes they performed them on the wrong finger ("The hardest part is not to make a mistake between the 2 fingers" [P2]). 7 participants mentioned slight proprioception problems as the cause of this inconsistency. Finally, neither the quantitative data nor the interview data reveal any significant effect or general preference for *Brightness* and *Size* enhancements. Only sparse personal preferences have been identified.

4.4 Resulting design recommendations

4.4.1 Relevance of the experiment design framework. Overall, participants found the representations both appealing and useful, which is encouraging for the widespread adoption of such representations. Most importantly, the tested simultaneous representations of microgesture efficiently convey the microgestures to be performed for a given set of commands. Please, keep in mind that we only considered two families. Consequently, although we did not observe any significant variability between these two families, results may vary from one family to another.

According to `Experiment 1`, **mixing families should be avoided (Recommendation 1, or Reco1).** It can create interaction effects, as we observed with the interpretation of the `tap` of the $_{MatchingShapes}$ family confused with the `hold` of the $_{ArrowAndBall}$ family. Furthermore, **`swipe` arrow strokes should not overlap (Reco2)** because when they do, arrows become ambiguous as to the direction of the `swipes` (see Figure 12a).

4.4.2 Representing `swipes`. P4 suggested to represent $_{A\&B}$ `swipes` by splitting the arrow strokes. This would make it easier to distinguish the directions, as they would not overlap at all (see Figure 12b). Even though they were not as explicit, other participants comments were consistent with this suggestion. They also agreed on the uselessness of redundant arrows. Figure 12c presents an example with one arrow instead of three. Furthermore, 9 out of 16 participants wanted to `swipe` the icon dedicated to representing the middle `tap` in the case of *Overlapping* representations. Their interpretation suggests that a unique command icon in the middle of the arrows strokes



(a) Original (b) Reworked (c) Removed redundant arrows

Fig. 12. Comparison of the (a) original and (b) reworked simultaneous representation of microgestures for the $\circlearrowright_{A\&B}$ family with or (c) without the redundant arrows[14].

could have been sufficient to convey opposite actions, e.g. volume up/down, associated to a `swipe` microgesture.

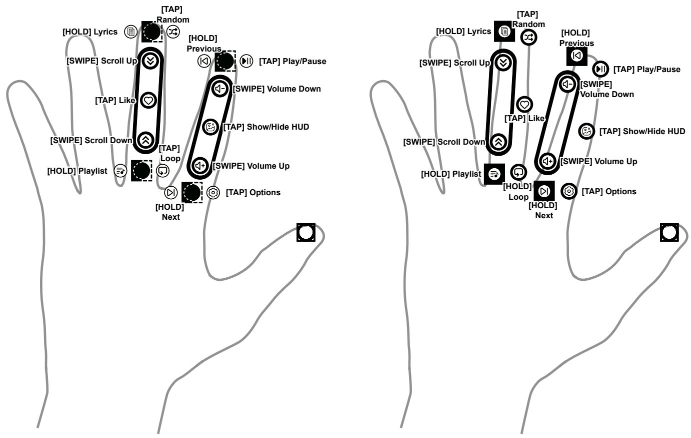
4.4.3 Key recommendations towards improved clarity. Considering the trade-off between density of information and clarity of simultaneous representations of microgestures, we expected that the participants would prefer the *Juxtaposition* strategy. The obtained results show a high overload for the *Overlapping* strategy, validating this intuition. Furthermore, the correlation between measures suggests that the overload should be minimized in order to improve the attractiveness and readability of the representation. Following this statement, ***Legends* should be avoided (Reco3)** as they are responsible for a significant increase in perceived overload. However, misplaced *Labels* tend to confuse users. An important number of persons naturally rely on the displayed position of the visual cues which can become a problem when *Labels* or command icons overlap the wrong zone or finger. Consequently, ***Juxtaposition* should be preferred to *Overlapping* as long as the command *Label* and icon positions do not confuse users (Reco4).**

`Experiment 1` involved 2 families. This allowed us to notice that the participants only recognized two levels of border width: tiny/no border and noticeably large border. Practitioners should thus **prefer a binary scale for size variations within the same application design (Reco5).**

4.4.4 User-dependant implications. In the result section, we indicated that the *Text* enhancement is perceived as responsible for a significant overload. Nevertheless, it is worth pointing out that 13 participants out of 32 found that it could help dispel any doubts about which microgesture to perform. The simultaneous representation of microgestures could thus offer a unique support for novice and intermediate users, which is investigated in `Experiment 2`.

Finally, we showed that preferences for one or other family led to Manichean remarks among the participants. In order to maximize the ease and attractiveness of their applications, practitioners should thus **leave the choice of the family to the user if possible (Reco6).**

[14]The reworked arrows are thinner with respect to the comments about the excessive thickness of the original arrows.



(a) Overlapping (b) Juxtaposition

Fig. 13. 14 Spotify commands: representation of the microgestures and commands with (a) *Overlapping* and (b) *Juxtaposition* strategies with the $\mathbf{O}_{MatchingShapes}$ family, *Labels* and the *Text* enhancement.

5 Experiment 2: A recommendation-based representation in practice

Practitioners now have insights on how to design simultaneous representations of microgestures. However, these representations have not been tested in context yet. We therefore explore this question in a second experiment.

5.1 Rationale of the simultaneous representation of microgestures tested

The most commonly used thumb-to-finger microgestures involve the `index` and `middle` fingers [13, 19, 30, 32] and 3 touch zones [12, 30]. Considering the encouraging results of `Experiment 1`, we thus choose an extreme case were `tap`, `swipe` and `hold` microgestures are simultaneously represented on both `index` and `middle` fingers. However, `Experiment 1` does not evaluate the case were 3 microgestures share the same zone of a finger. Aware of this limit, we consider a total of 14 microgestures: 6 `taps`, 4 `holds` and 4 `swipes` (see Figure 13). As previously done in [5, 7], we choose a musical application, here Spotify[15], and select 14 commands[16]. Following Reco1, Reco2 and Reco3, we use the $\mathbf{O}_{MatchingShapes}$ family with a *Labels*-based integration of the command names. Figure 13 depicts the resulting representations using *Juxtaposition* and *Overlapping* strategies. We should avoid *Labels* in the inter-digital space as they tend to confuse users (Reco4). Consequently, we use the *Overlapping* representation (see Figure 13a) in this experiment.

5.2 Participants & Apparatus

We recruited 14 unpaid volunteers ranging in age from 19 to 30 years old (M=25, σ=3). 5 self-identified as women, 7 as men, 1 as non-binary and 1 as agender. 5 were already familiar with Spotify. They covered a wide range of professions, from midwifes to engineers and drama students. They were recruited via social networks and on the local campus. Considering the results of `Experiment 1`, we added a *Text* enhancement (displaying the type of the microgesture in addition to the name of the command) and limited this study to users who had not participated in

[15]Spotify website: www.spotify.com
[16]Aware of color-blind users, we choose black and white icons.

`Experiment 1`, to ensure that all participants had a novice level of expertise in microgestures and their representations.

We displayed the simultaneous representation of 14 microgestures on a Microsoft Surface 9. The representation was sized to be displayed on a 8-inch smartphone. The screen of the experimenter's smartphone was also broadcast live with the Spotify application running (see Appendix C).

5.3 Procedure

On-hand devices such as gloves or trackers could impair the participant's motion and provide clues as to the types of gestures recognized. After technical tests with Mediapipe[17], a vision-based solution which does not require to equip users, we concluded it was not reliable enough to be used in an experiment involving both the index and the middle fingers with their 3 phalanxes. Hence, we conducted a Wizard of Oz experiment, simulating vision-based microgesture recognition system. Inspired by our tests with Mediapipe, we asked participants to place their hands correctly in front of the webcam and to exaggerate microgestures to "avoid detection errors". The application displayed a screen-share from the experimenter's phone, unbeknownst to participants. In the extremely rare cases[18] where the experimenter was unsure of the performed microgesture, they waited for the participants to repeat the microgesture.

The experiment was based on the following procedure, which took around 15 minutes to complete:

(1) Consent form and preliminary questions
(2) Free exploration phase (3 minutes maximum)
(3) Scenarios execution (all performed once)
  (a) **Like** two target songs among a playlist of 12 songs
  (b) Increase the **Volume** of the music, then decrease it at ease
  (c) Activate **Random** mode or **Loop** mode
  (d) Get the last sentence of the **Lyrics** of the target song
(4) Evaluation of the representation with a 5-point Likert scale[19] on 4 topics inspired by Yu *et al.*' statements [60]: convenience, easiness, pleasure of use, i.e. fun, and likelihood of wanting this type of help for future microgesture-based applications.
(5) Revealing the wizard

With only the graphical representation of microgestures, 8 participants did not understand that we were focusing on thumb-to-finger microgestures. For those, we gave as a hint that the considered microgestures involved "a contact between the thumb and other fingers" during the free exploration phase.

5.4 Results

All participants believed that their microgestures were automatically recognized ("I did not see a thing!" [P9], "turning on the webcam is very useful [for boosting confidence in the detection system]" [P15]). Overall, the participants were satisfied with the simultaneous representation of microgestures ("having a chart like this is clearly useful" [P8]).

5.4.1 Usability and appeal of the representation based on the reported measures. The simultaneous representation of microgestures has been judged convenient by all users except 1 (Mdn=0.75, CI=[0.75, 0.88], M=0.79, CI=[0.68, 0.88]). Users considered the system as fun (Mdn=1, CI=[0.75,

[17]Mediapipe website: https://developers.google.com/mediapipe
[18]It occurred less than 20 times out of at least 400 microgestures performed while completing the scenarios (orders of magnitude).
[19]We encoded the Likert scale from *Strongly disagree* to *Strongly agree* using multiples of 0.25 between 0 and 1 inclusive as done in `Experiment 1`.

1], M=0.82, CI=[0.7, 0.93]) but not so easy, even if accessible (Mdn=0.62, CI=[0.5, 0.75], M=0.59, CI=[0.46, 0.7]). Lastly, although they were unable to make comparisons with other help systems, the participants were positive about the idea of having this kind of help available for future microgesture-based applications (Mdn=1, CI=[0.75, 1], M=0.89, CI=[0.82, 0.95]). 3 participants out of 14 were confused by the swipe direction. All participants left *Text* enhancement active throughout the experiment, which is consistent with the results of `Experiment 1`. See Appendix C for the full distributions of participants responses.

5.4.2 Initial learning of microgestures. The difference between easiness and convenience has been explained by some participants by a "steep learning curve" [P6] ("once you've understood, it's convenient" [P1, P5, P13], "not so easy but convenient" [P8, P10]). 10 participants out of 14 were initially confused by the term “microgesture” and made extremely subtle flexes ($_{i}\overset{\circ\circ}{\blacktriangle}$ | $_{i}\overset{\circ\circ}{\blacktriangledown}$) during the free exploration phase. 2 of them figured out on their own that they had to use their thumbs, but the 8 others were completely lost and quickly ran out of ideas.

5.4.3 Other design comments. The participants were mostly positive about the help system ("the idea is really interesting" [P5], "it's fire! [...] I cannot imagine a system more explicit than this one" [P9], "if you are having trouble getting started, or if you have forgotten, you can have a cheatsheet with all the gestures" [P11]). However, P6, P7 and P13 explained that the drawn hand itself could be a source of confusion because, during interaction, they saw the back of their hand and not their palm. They suggested to add lines representing the wrinkles of the palm to avoid the confusion. P2 and P6 would have preferred a representation containing less information, which is consistent with the results of `Experiment 1` ("too many things at the same place" [P2], "quite stacked but at least everything was here" [P6]).

6 Experiment 3: Evaluating user preferences according to screen size

`Experiment 2` shows that simultaneous representations of microgestures can be easily used and understood, with a size suitable for display on an 8-inch smartphone. In a third experiment, we now evaluate their strengths and weaknesses against a baseline with smaller wearable devices such as smaller smartphones and smartwatches. The baseline consists of a mosaic of representations for the same microgestures and associated commands as in `Experiment 2`. For representing microgestures, we use $\circlearrowright_{ArrowAndBall}$, the best design family identified in [35]. Within the mosaic, we spatially group representations by finger, by zone and by microgesture type (see Figure 14b). For each screen size condition (smartphone and smartwatch), we maximize the size of the simultaneous representation of microgestures as well as of the baseline, in order to make optimum use of each representation[20].

6.1 Participants & Apparatus

We recruited 24 participants (12 self-identified as men, 12 as women) ranging in age from 19 to 32 (M=24, sigma=2.9), none of whom had participated in the previous experiments. They covered a wide range of professions, from nurses to engineers to event professionals. They were recruited via social networks and on the local campus. 6 owned a smartwatch.

We used a Samsung Galaxy A13 (6.6 inches, PPI=400). We attached the same smartphone to the wrist using an armband to simulate the context of a smartwatch, as shown in Figure 14. This setup

[20]The resulting icons and font sizes were 1.134 times larger in the simultaneous representation of microgestures than in the baseline for the smartphone screen size (~1mm for one uppercase character height in both cases) and 1.259 times larger for the smartwatch screen size (~0.5mm for one uppercase character height in both cases).

has already been used in experiments [43, 51]. This guarantees exactly the same PPI and brightness conditions for both screen sizes (smartphone and smartwatch).

6.2 Procedure

We considered 2 binary variables: the type of representation and the screen size. It resulted in 4 conditions (*Simultaneous-Smartphone*, *Baseline-Smartphone*, *Simultaneous-Smartwatch* and *Baseline-Smartwatch*). We tested all 24 possible orders. The experiment was based on the following procedure, which took about 15 minutes to complete.

(1) Introductory video, consent form and preliminary questions
(2) For each of the 4 conditions
   (a) For 3 microgestures (once for each type of microgesture): given the command, find the associated microgesture
   (b) Evaluation of the easiness to find the commands and understand the microgestures (5-point Likert scale[21])
(3) User preferences between the two representations (*Simultaneous* or *Baseline*) according to screen size
(4) Open questions to fuel discussion
   (a) What would be the ideal number of microgestures for a given type of representation and device?
   (b) What do you think of sequential display (time multiplexing) of microgestures, like slideshows, instead of static display of a mosaic of representations?

6.3 Results

The simultaneous representation of microgestures and the baseline both achieve encouraging results in terms of ease of finding commands (Mdn=0.75, CI=[0.62, 0.88] for both (N=24)) and understanding the associated microgestures (Mdn=0.88, CI=[0.62, 0.88] for *Simultaneous* (N=24) ;

[21]We encoded the Likert scale from *Strongly disagree* to *Strongly agree* using multiples of 0.25 between 0 and 1 inclusive as in `Experiment 1`.

(e) Experiment setup

Fig. 14. Experiment conditions: (a) the simultaneous representation of microgestures displayed on a smartphone or (b) the baseline displayed on the same smartphone and (c) the simultaneous representation of microgestures displayed on a smartwatch or (d) the baseline displayed on the same smartwatch. The smartwatch was (e) simulated by a smartphone

Mdn=0.88, CI=[0.75, 1] for *Baseline* (N=24))[22]. Only P4 felt heavily confused by the simultaneous representation of microgestures.

6.3.1 Impact of screen size. With the smartphone condition, most participants did not express a clear preference ("both are good" [P9, P21]). However, we deliberately formulated our questions in such a way as to force them to choose one or the other, in order to guarantee discussion. 13 participants out of 24 preferred the baseline because it was easier to understand ("[it's] more obvious" [P0], "the mosaic speaks for itself" [P3]). 11 participants out of 24 preferred the simultaneous representation of microgestures, which they found "clearer because more condensed" [P2, P7] and more "intuitive" [P12, P15, P17].

With the smartwatch condition, 20 participants out of 24 considered the simultaneous representation of microgestures better than the baseline. For 16 of them, the images and texts in the baseline were "too small" [P7, P13, P21] due to the number of microgestures represented ("the more the microgestures there are, the more the mosaic becomes hard to understand" [P10]). In addition, P2, P7 and P9 insisted that we could have further optimized the font size of the simultaneous representation of microgestures by further cropping the wrist. Nevertheless, we did not observe a significant difference between both types of representation in terms of ease of finding commands (Mdn=0.75, CI=[0.74, 0.75] for *Simultaneous* (N=24) ; Mdn=0.62, CI=[0.5, 0.75] for *Baseline* (N=24)) and of understanding the associated microgestures (Mdn=1, CI=[0.5, 1] for *Simultaneous* (N=24) ; Mdn=0.88, CI=[0.75, 1] for *Baseline* (N=24)).

6.3.2 Perceived strengths and weaknesses. Participants often justified their preferences with utilitarian arguments. The simultaneous representation of microgestures required an initial effort to be understood ("I needed more time at first glance" [P10, P13, P19], "at first, I did not understand but after a few seconds I did" [P7, P14]). However, the spatial organization of the commands seemed clearer with this representation than with the baseline ("you can find the information and the gesture more quickly" [P12], "the organization is easier to understand" [P0, P7], "the [simultaneous representation] is preferable for knowing where to make the gesture" [P11]). Consequently, the baseline seems to be better for initial learning ("good for a quick tutorial" [P1]) whereas the simultaneous representation of microgestures seems to be better for command recall ("once you want something handy, [the simultaneous representation] is better" [P1], "to recall, [the simultaneous representation] is perfect" [P5, P14]).

In addition, participants highlighted that the simultaneous representation of microgestures seemed more adapted than the baseline to keep focus on a primary task ("I feel like I can see it and do [the microgesture] right away whereas with the [baseline] I need to focus on my fingers" [P13], "I would have preferred the [simultaneous representation of microgestures] with any game-related context" [P0]). P5, P13 and P15, explained it by being more able to map the commands onto their hand since the representation only uses one condensed hand image.

6.3.3 Feedback on open questions. 12 participants out of 24 were highly skeptical about a solution based on temporal multiplexing ("time-consuming" [P14, P19], "I don't think I would like it" [P18, P22, P23]). In addition, 12 participants considered that the baseline should represent 5 or 7 microgestures at most for comfortable use on a smartwatch screen.

[22]For further details, see Appendix D, Figure 20 and Figure 21

7 Discussion

We first conducted an experiment to compare several design strategies (`Experiment 1`). We used the resulting recommendations to create a simultaneous representation of microgestures. We then evaluated its usability in context (`Experiment 2`) and its suitability for small screen sizes compared with a baseline made of a mosaic of single-picture representations (`Experiment 3`). To begin the general discussion, we can supplement the recommendations from `Experiment 1` with the results of `Experiment 2` and `Experiment 3`.

7.1 Enriched design recommendations

Considering the procedures and verbatims of `Experiment 2` and `Experiment 3`, we can confidently state that, with a *Text* enhancement, simultaneous representations of microgestures are satisfactory for novice users[23]. We therefore recommend **displaying the microgesture type next to the command name for novice users and not for intermediate users**[24] **(Reco7)**. However, novice users would likely need a quick tutorial to introduce the principle of thumb-to-finger microgestures and avoid confusion.

Representing all the microgestures on a single image of the hand makes it possible to understand the spatial and semantic organization of commands. It is then easier to find a similar command to one already found, and to locate the exact finger zone on a small screen. Consequently, simultaneous representations of microgestures appear **especially promising for command recall**. In addition, practitioners need to take into account the screen real estate when making their design choices [18]. Their design solutions should easily accommodate a variable number of microgestures. As a result, simultaneous representations of microgestures seem particularly well suited to smartwatches compared with other representations. Nevertheless, they remain suitable for larger display spaces.

Our results thus suggest that, compared to a mosaic of single-picture representations, simultaneous representations of microgestures should be **preferred for intermediate and expert users in any case and for novice users if displayed on a smartwatch (Reco8)**. If the help is not displayed on a smartwatch, and particularly if it is only intended to be used for initial learning, practitioners should consider using a mosaic of single-picture representations.

7.2 Actionable synthesis

We have compiled all these key ideas into a ready-to-use decision tree (see Figure 15) to *quickly* guide designers towards a representation that *avoids any ambiguity* and *minimizes the visual overloading*. We deeply encourage practitioners to use it, to enrich it with their own results and to share it with their colleagues.

7.3 Practical applications

7.3.1 Designing a smartwatch application. Nat develops a hiking application for smartwatches. Users can track their journey on the map, set or receive emergency alerts, e.g. falls, weather disasters, and specify points of interest. Nat knows that some functionalities like navigating on the map are not easy to use. She heard about Meta and Apple's research [1, 2] on microgestures and finds it promising. Nat prototypes microgesture interactions for her application with a custom glove. She is delighted of the result and lends her prototype to her friend John to make him try it. However, he is not familiar with smartwatches and microgestures at all. As a result, he ends up trying random gestures, getting very frustrated.

[23]The *wow effect* [34] that could have played a role in the results obtained in `Experiment 2` is strongly mitigated in `Experiment 3` since participants were exposed to different types of representation.
[24]Refer to the results of `Experiment 1`

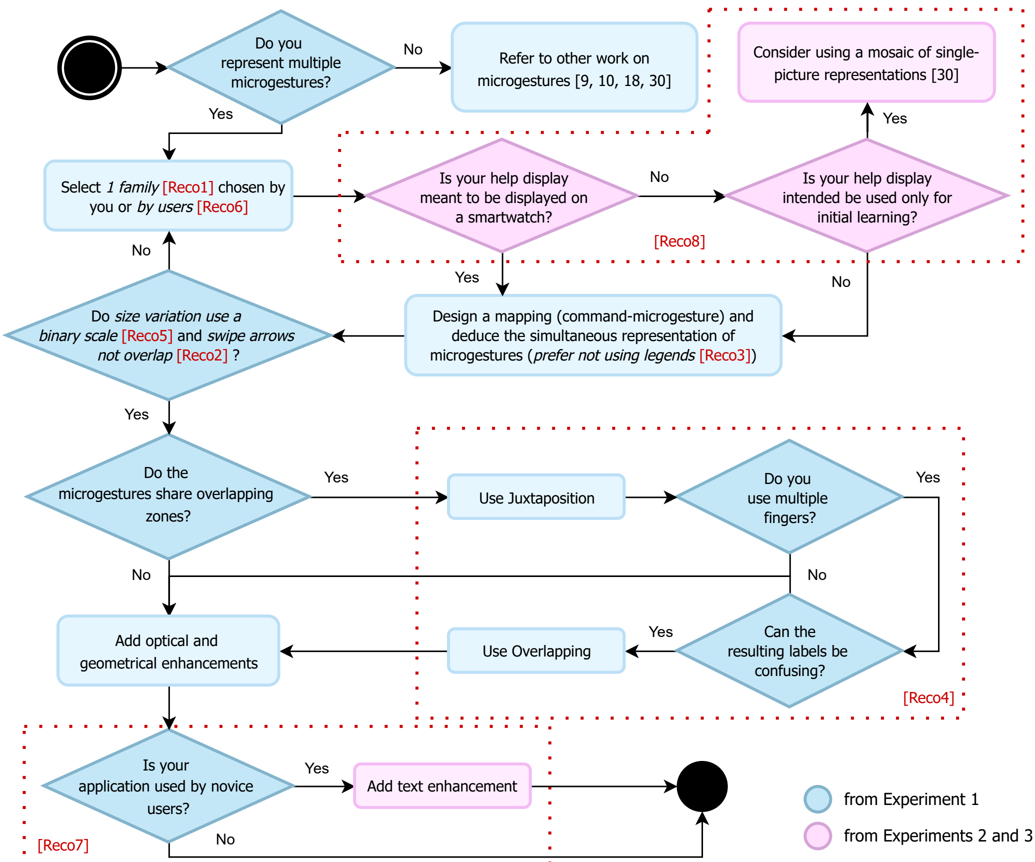


Fig. 15. Decision tree made for practitioners to inform the design of their applications. Among others, you can refer to [35] for other work on the representation of microgestures.

Nat turns to literature to help her friend learn the commands of her application and quickly discovers our decision tree. Since she is working with a smartwatch, Nat decides to create her own simultaneous representation of microgestures (see Reco8). She chooses a family (see Reco1) and designs a mapping for her application (see Reco3). Considering that John is a novice, she specifies the type of microgesture to perform for each command (see Reco7). After a few seconds, John grasps how the representation works. With a few tries, he understands that with swipes, he can navigate and zoom on the map, and with taps, he can specify points of interest without even looking at the screen. He does not seem frustrated and can find a forgotten command easily[25].

7.3.2 Designing a multi-devices application. During her research, Nat discovered that microgestures would be perfect for interacting with AR glasses [49]. As a result, Nat conceives an AR version of her hiking application with additional features. For each new feature, she associates a new microgesture-based command. John would be able to highlight the right path to follow or get environment-based information, e.g. tree species or mountain names, with holds.

[25]Please note that any hypothesis about **learning** a set of microgestures needs to be confirmed by future work (see 7.5)

Nat believes that using the same type of help display should ease the transition from smartwatch to AR glasses. Furthermore, the help is not only intended to be used for initial learning but also for command recall. Thus, Nat creates another simultaneous representation of microgestures to display the commands that are specific to the AR version (see Reco8). She follows the same steps than before and then asks John to test the AR glasses app. As he already tried the smartwatch version, he quickly understands that the core commands are the same. When he invokes the help, he notices the second representation. After a few seconds to understand the new representation, he smoothly learns and retains the new commands[25].

7.4 Limitations

Our work gives no indication of how to represent flexes (▲| ▼), draws (↻; ⋦; etc.) and grasp-based microgestures (■(✹)). Furthermore, our study considers a maximum of two microgestures on the same finger zone and evaluates only two design families. We have not considered the ring and pinky fingers, nor the palm and nails.

Simultaneous representations of microgestures themselves suffer limitations. First, they depend on the design family but also on the chosen microgesture types, the number of microgestures and their location on the hand. Second, participants were often confused at first glance. Simultaneous representations of microgestures thus require us to spend a few seconds or tens of seconds overcoming their apparent complexity.

7.5 Future work

Benefits for space-efficient encoding depends on the microgestures depicted (number, type, zones) and their designed representations. Future studies should deepen our understanding of the underlying trade-offs, especially for grasp-based microgestures that completely change the context of use.

Learning a microgesture set is undoubtedly impacted by the visual aids provided. The verbatim we have collected gives indications that deserve further investigation to establish the strengths and weaknesses of simultaneous representations of microgestures for learning a set of microgestures.

A heuristic evaluation of the recommendations themselves should be conducted with designers to further inform on their practical effectiveness. A study could be conducted with two groups of designers. The first group would be informed of the recommendations whereas the other would not. In this type of A-B comparison, users would evaluate the efficiency of the representations developed with and without the recommendations in mind. The results could also potentially enrich the set of recommendations.

Combining spatial and temporal multiplexing could be an interesting way of finding a compromise between information density and clarity for simultaneous representations of microgestures. The same amount of information could be distributed between the fingers and/or presented sequentially by the help system.

Unforeseen practical applications were discovered thanks to this study. 3 participants of `Experiment 1` had dyslexia and 2 of them noticed similarities with the exercises they had to do to compensate their difficulties, i.e. find the right word from a set of close spellings which causes confusion. Furthermore, two participants were nurses and mentioned that serious games could be designed to guide patients during hand rehabilitation. Two participants of `Experiment 2` expressed their interest in this type of games as patients. The first had an injured hand and the other suffered side-effects from a stroke. Collaborative work could be carried out with speech therapists and physiotherapists to design effective exercises based on the simultaneous representation of microgestures. Another unforeseen application is underwater interaction. In `Experiment 3`, P5 explicitly mentioned that their smartwatch's touch controls do not work underwater. Consequently, they would be interested in microgestures to control their smartwatch in the pool or shower, which would involve using simultaneous representations of microgestures for command recall.

8 Conclusion

In this paper, we conducted 3 experiments to identify and evaluate several aspects of the simultaneous representation of microgestures. Given the current state of knowledge, this work provides an essential starting point for the creation of interaction illustrations suitable for all screen sizes. To help practitioners design simultaneous representations of microgestures, we propose 8 design recommendations contextualized using a decision tree. We also provide Inkscape plugins to easily create these representations. We hope these contributions will encourage further studies on how to represent microgestures and commands in applications. This is crucial for the widespread adoption of microgesture-based interaction: users need to know which microgestures are available and which commands are associated with them. Finally, the discussion expands the scope of application initially imagined for this type of representation and proposes new contexts for using microgestures.

Acknowledgments

This work was partly supported by the French National Research Agency (ANR) project MIC (ANR-22-CE33-0017).

References

[1] Inside Facebook Reality Labs: Wrist-based interaction for the next computing platform, March 2021.

[2] Apple Watch User Guide, 2024.

[3] Axel Antoine, Sylvain Malacria, Nicolai Marquardt, and Géry Casiez. Interaction Illustration Taxonomy: Classification of Styles and Techniques for Visually Representing Interaction Scenarios. In *CHI 2021 - ACM Conference on Human Factors in Computing Systems*, CHI '21, page 22, Yokohama, Japan, May 2021. Association for Computing Machinery.

[4] Apple. Trackpad gestures for iPad, 2022.

[5] Bijan Aryana and Torkil Clemmensen. Mobile Usability: Experiences From Iran and Turkey. *International Journal of Human-Computer Interaction*, 29(4):220–242, March 2013.

[6] Jacques Bertin. *Semiology of Graphics: Diagrams, Networks, Maps.* Esri Press, USA, 2011.

[7] Roger Boldu, Alexandru Dancu, Denys Matthies, Pablo Cascon, Shanaka Ransir, and Suranga Nanayakkara. Thumb-In-Motion: Evaluating Thumb-to-Ring Microgestures for Athletic Activity. In *Proceedings of the Symposium on Spatial User Interaction*, SUI '18, pages 150–157, New York, NY, USA, October 2018. Association for Computing Machinery.

[8] Rita Borgo, Min Chen, Markus Höferlin, Kuno Kurzhals, Phil Legg, Simon Walton, and Daniel Weiskopf. EG2013 Tutorial on Video Visualization.

[9] Manuel Bouillon and Eric Anquetil. Online active supervision of an evolving classifier for customized-gesture-command learning. *Neurocomputing*, 262:77–89, November 2017.

[10] J.M. Carey. *Human Factors in Information Systems: An Organizational Perspective.* Ablex Series in Computational Science. Ablex Publishing Corporation, 1991.

[11] Adrien Chaffangeon, Alix Goguey, and Laurence Nigay. μGlyphe: une Notation Graphique pour Décrire les Microgestes. In *33ème conférence internationale francophone sur l'Interaction Humain-Machine (IHM'22)*, IHM '22, page 13, New York, NY, USA, April 2022. Association for Computing Machinery.

[12] Adrien Chaffangeon Caillet, Alix Goguey, and Laurence Nigay. μGlyph: a Microgesture Notation. CHI '23, page 28, New York, NY, USA, 2023. Association for Computing Machinery.

[13] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. User Elicitation on Single-hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3403–3414, San Jose California USA, May 2016. ACM.

[14] Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, Kai-Yin Cheng, Chao-Huai Su, Mike Y. Chen, Wen-Huang Cheng, and Bing-Yu Chen. FingerPad: private and subtle interaction using fingertips. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 255–260, St. Andrews Scotland, United Kingdom, October 2013. ACM.

[15] Chaomei Chen. Information visualization: Information visualization. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):387–403, July 2010.

[16] M. Chen, S. Walton, K. Berger, J. Thiyagalingam, B. Duffy, H. Fang, C. Holloway, and A. E. Trefethen. Visual Multiplexing. In *EuroVis '14: Proceedings of the 16th Eurographics Conference on Visualization*, volume 33, pages 241–250, Swansea, UK, June 2014. Eurographics Association.

[17] Victor Chen, Xuhai Xu, Richard Li, Yuanchun Shi, Shwetak Patel, and Yuntao Wang. Understanding the Design Space of Mouth Microgestures, June 2021. arXiv:2106.00931 [cs].

[18] William Delamare, Daichi Harada, Luxi Yang, and Xiangshi Ren. Guiding gaze gestures on smartwatches: Introducing fireworks. *International Journal of Human-Computer Studies*, 183:103196, March 2024.

[19] Bastian Dewitz, Frank Steinicke, and Christian Geiger. Functional Workspace for One-Handed Tap and Swipe Microgestures. In *Mensch und Computer 2019 - Workshopband*. Gesellschaft für Informatik e.V., 2019. Publisher: Gesellschaft für Informatik e.V.

[20] Pierre Dragicevic. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*, pages 291–330. Springer International Publishing, Cham, 2016. Series Title: Human–Computer Interaction Series.

[21] Jason Dykes, Jo Wood, and Aidan Slingsby. Rethinking Map Legends with Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):890–899, November 2010.

[22] Shariff Am Faleel, Michael Gammon, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. HPUI: Hand Proximate User Interfaces for One-Handed Interactions on Head Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4215–4225, November 2021.

[23] Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris, and Daniel Wigdor. ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '09*, page 165, Banff, Alberta, Canada, 2009. ACM Press.

[24] Koumei Fukahori, Daisuke Sakamoto, and Takeo Igarashi. Exploring Subtle Foot Plantar-based Gestures with Sock-placed Pressure Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3019–3028, Seoul Republic of Korea, April 2015. ACM.

[25] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, October 2011.

[26] I. Golebiowska. Strategies of legend and map integration. In *Proceedings of the 25th International Cartographic Conference Paris, 3-8 July 2011*. ICC, Paris, 2011. OCLC: 781048687.

[27] Tovi Grossman, Pierre Dragicevic, and Ravin Balakrishnan. Strategies for accelerating on-line learning of hotkeys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1591–1600, San Jose California USA, April 2007. ACM.

[28] Charles D. Hansen, Min Chen, Christopher R. Johnson, Arie E. Kaufman, and Hans Hagen, editors. *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*. Mathematics and Visualization. Springer London, London, 2014.

[29] Chris Harrison, Desney Tan, and Dan Morris. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–462, Atlanta Georgia USA, April 2010. ACM.

[30] Da-Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong-Hao Liang, De-Nian Yang, Yi-Ping Hung, and Bing-Yu Chen. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1526–1537, San Jose California USA, May 2016. ACM.

[31] Waqas Javed and Niklas Elmqvist. Exploring the design space of composite visualization. In *2012 IEEE Pacific Visualization Symposium*, pages 1–8, Songdo, Korea (South), February 2012. IEEE.

[32] Haiyan Jiang, Dongdong Weng, Zhenliang Zhang, and Feng Chen. HiFinger: One-Handed Text Entry Technique for Virtual Environments Based on Touches between Fingers. *Sensors*, 19(14):3063, July 2019. Number: 14.

[33] Anjali Khurana, Michael Glueck, and Parmit K. Chilana. Do I Just Tap My Headset?: How Novice Users Discover Gestural Interactions with Consumer Augmented Reality Applications. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(4):1–28, December 2023.

[34] Marion Koelle, Thomas Olsson, Robb Mitchell, Julie Williamson, and Susanne Boll. What is (un)acceptable?: thoughts on social acceptability in HCI research. *Interactions*, 26(3):36–40, April 2019.

[35] Vincent Lambert, Adrien Chaffangeon Caillet, Alix Goguey, Sylvain Malacria, and Laurence Nigay. Studying the Visual Representation of Microgestures. In *Proceedings of the 25th international conference on Human-computer interaction with mobile devices & services - MobileHCI '23*, volume 7, Athens, Greece, 2023. ACM Press.

[36] Guangchuan Li, David Rempel, Yue Liu, Weitao Song, and Carisa Harris Adamson. Design of 3D Microgestures for Commands in Virtual Reality or Augmented Reality. *Applied Sciences*, 11(14):6375, January 2021. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

[37] Christian Loclair and Sean Gustafson. PinchWatch: A Wearable Device for One-Handed Microinteractions, 2010.

[38] Sehi LYi, Jaemin Jo, and Jinwook Seo. Comparative Layouts Revisited: Design Space, Guidelines, and Future Directions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1525–1535, February 2021.

[39] Eva Mackamul, Géry Casiez, and Sylvain Malacria. Exploring Visual Signifier Characteristics to Improve the Perception of Affordances of In-Place Touch Inputs. *Proceedings of the ACM on Human-Computer Interaction*, 7(MHCI):1–32, September 2023.

[40] Eva Mackamul, Géry Casiez, and Sylvain Malacria. Clarifying and differentiating discoverability. *Human–Computer Interaction*, pages 1–26, 2024.

[41] Nathan Magrofuoco, Jorge-Luis Perez-Medina, Paolo Roselli, Jean Vanderdonckt, and Santiago Villarreal. Eliciting Contact-Based and Contactless Gestures With Radar-Based Sensors. *IEEE Access*, 7:176982–176997, 2019.

[42] Francis T. Marchese and Ebad Banissi, editors. *Knowledge Visualization Currents: From Text to Art to Culture*. Springer London, London, 2013.

[43] Stephen Oney, Chris Harrison, Amy Ogan, and Jason Wiese. ZoomBoard: a diminutive qwerty soft keyboard using iterative zooming for ultra-small devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2799–2802, Paris France, April 2013. ACM.

[44] Thomas Ostermann and Dennis Anheyer. "Don't You Understand? It's an Effect of $d = 0.09$ Between the Groups, Stupid!" Using Common Language Effect Sizes to Improve the Understandability of Study Results. *Journal of Integrative and Complementary Medicine*, 29(8):452–454, August 2023.

[45] Francisco Perella-Holfeld, Shariff Am Faleel, and Pourang Irani. Evaluating design guidelines for hand proximate user interfaces. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pages 1159–1173, Pittsburgh PA USA, July 2023. ACM.

[46] Manuel Prätorius, Dimitar Valkov, Ulrich Burgbacher, and Klaus Hinrichs. DigiTap: an eyes-free VR/AR symbolic input device. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*, pages 9–18, Edinburgh, Scotland, 2014. ACM Press.

[47] Annemarie Quispel, Alfons Maes, and Joost Schilperoord. Aesthetics and Clarity in Information Visualization: The Designer's Perspective. *Arts*, 7(4):72, November 2018.

[48] Judy Robertson and Maurits Kaptein, editors. *Modern Statistical Methods for HCI*. Human–Computer Interaction Series. Springer International Publishing, Cham, 2016.

[49] Shardul Sapkota, Ashwin Ram, and Shengdong Zhao. Ubiquitous Interactions for Heads-Up Computing: Understanding Users' Preferences for Subtle Interaction Techniques in Everyday Settings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, MobileHCI '21, pages 1–15, New York, NY, USA, September 2021. Association for Computing Machinery.

[50] Adwait Sharma, Michael A. Hedderich, Divyanshu Bhardwaj, Bruno Fruchard, Jess McIntosh, Aditya Shekhar Nittala, Dietrich Klakow, Daniel Ashbrook, and Jürgen Steimle. SoloFinger: Robust Microgestures while Grasping Everyday Objects. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, Yokohama Japan, May 2021. ACM.

[51] Muhammad Shoaib, Stephan Bosch, Ozlem Incel, Hans Scholten, and Paul Havinga. Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors. *Sensors*, 16(4):426, March 2016.

[52] Mohamed Soliman, Franziska Mueller, Lena Hegemann, Joan Sol Roo, Christian Theobalt, and Jürgen Steimle. FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ISS '18, pages 177–187, New York, NY, USA, November 2018. Association for Computing Machinery.

[53] Maxim Spur, Vincent Tourre, Erwan David, Guillaume Moreau, and Patrick Callet. MapStack: Exploring Multilayered Geospatial Data in Virtual Reality:. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 88–99, Valletta, Malta, 2020. SCITEPRESS - Science and Technology Publications.

[54] Hsin-Ruey Tsai, Te-Yen Wu, Da-Yuan Huang, Min-Chieh Hsiu, Jui-Chun Hsiao, Yi-Ping Hung, Mike Y. Chen, and Bing-Yu Chen. SegTouch: Enhancing Touch Input While Providing Touch Gestures on Screens Using Thumb-To-Index-Finger Gestures. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2164–2171, Denver Colorado USA, May 2017. ACM.

[55] András Vargha, Harold D. Delaney, and Andras Vargha. A Critique and Improvement of the "CL" Common Language Effect Size Statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25(2):101, 2000.

[56] Craig Villamor, Dan Willis, and Luke Wroblewski. Touch Gesture Reference Guide, 2010.

[57] Jérémy Wambecke, Alix Goguey, Laurence Nigay, Lauren Dargent, Daniel Hauret, Stéphanie Lafon, and Jean-Samuel Louis de Visme. M[eye]cro: Eye-gaze+Microgestures for Multitasking and Interruptions. *Proceedings of the ACM on Human-Computer Interaction*, 5(EICS):1–22, May 2021. Number: EICS.

[58] Eric Whitmire, Mohit Jain, Divye Jain, Greg Nelson, Ravi Karkar, Shwetak Patel, and Mayank Goel. DigiTouch: Reconfigurable Thumb-to-Finger Input and Text Entry on Head-mounted Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–21, September 2017. Number: 3.

[59] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. A Taxonomy of Microinteractions: Defining Microgestures Based on Ergonomic and Scenario-Dependent Requirements. In *Human-Computer Interaction – INTERACT 2011*, volume 6946, pages 559–575. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. Series Title: Lecture Notes in Computer Science.

[60] Chun Yu, Xiaoying Wei, Shubh Vachher, Yue Qin, Chen Liang, Yueting Weng, Yizheng Gu, and Yuanchun Shi. HandSee: Enabling Full Hand Interaction on Smartphone with Front Camera-based Stereo Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

A Task success according to the commands

4200 representations were implemented in `Experiment 1` to ensure a sufficient randomness among the 4 conditions and the possible layouts for their corresponding representations and commands. Consequently, not all commands have been seen by our participants which is a common point with Grossman *et al.*' study [27] from which our command set have been taken. Figure 16 shows that, for an overwhelming majority, they were not any error due to a confusion between icons. Even if, on the whole, this is not a problem, we noticed that 15 out of 32 participants showed real difficulty in distinguishing between Baseball and Basketball. This can be explained by the close spelling of the two commands, despite their very different icons.

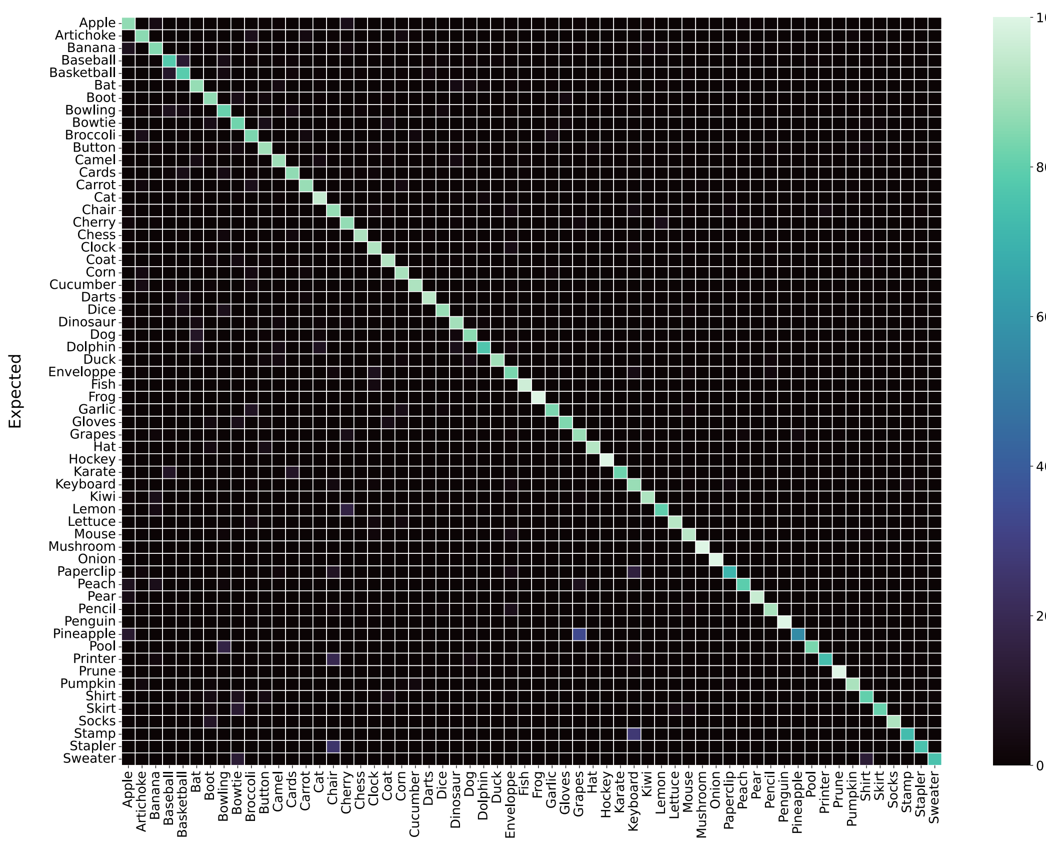


Fig. 16. Performed command depending on the expected command (all groups aggregated)

B Additional measure tables and figures

`Experiment 1` used 5 measures to evaluate each representation : success, easiness, overload, appeal and readability. Table 2a and Table 2b refer to success and overload respectively and are referenced in the body of this paper. Table 3a, Table 3b and Table 3c refer to easiness, appeal and readability respectively and are referenced here to help with reading.

Finally, Figure 17 depicts the correlation measures referenced in subsubsection 4.3.6.

		Microgestures combo	Tap-Swipe				Tap-Hold			
		Spatial strategy	Overlapping		Juxtaposition		Overlapping		Juxtaposition	
		Family	A&B	MaS	A&B	MaS	A&B	MaS	A&B	MaS
Finger combo	Enhancement	Integration								
Index	Default	Labels	1.0	0.95	0.95	0.98	0.83	0.73	0.98	0.98
		Legend	0.98	1.0	1.0	0.98	0.75	0.73	1.0	1.0
	Size	Labels	0.92	1.0	0.95	0.95				
		Legend	0.95	0.98	0.98	1.0				
	Brightness	Labels	0.95	0.98	1.0	0.95	0.75	0.75	1.0	1.0
		Legend	0.98	0.98	0.95	0.98	0.71	0.81	0.98	0.88
	Text	Labels	1.0	1.0	1.0	0.92	0.98	1.0	1.0	1.0
		Legend	0.98	0.92	0.92	0.98	1.0	0.98	0.98	1.0
Index-Middle	Default	Labels	0.98	0.99	0.98	0.96	0.47	0.7	0.99	0.9
		Legend	0.92	0.96	0.98	0.94	0.64	0.58	0.97	0.92
	Size	Labels	0.96	1.0	0.9	0.98				
		Legend	0.88	0.96	0.98	0.96				
	Brightness	Labels	0.92	0.98	0.99	0.98	0.89	0.81	0.96	0.92
		Legend	0.88	0.99	0.94	0.95	0.74	0.62	0.86	0.81
	Text	Labels	0.94	0.99	0.98	1.0	0.69	0.7	0.85	0.85
		Legend	0.94	0.94	1.0	0.82	0.46	0.41	0.97	0.86

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

(a) **Easiness**

		Microgestures combo	Tap-Swipe				Tap-Hold			
		Spatial strategy	Overlapping		Juxtaposition		Overlapping		Juxtaposition	
		Family	A&B	MaS	A&B	MaS	A&B	MaS	A&B	MaS
Finger combo	Enhancement	Integration								
Index	Default	Labels	0.69	0.81	0.75	0.84	0.72	0.78	0.78	0.81
		Legend	0.75	0.72	0.78	0.84	0.59	0.59	0.69	0.62
	Size	Labels	0.69	0.78	0.84	0.78				
		Legend	0.75	0.69	0.59	0.72				
	Brightness	Labels	0.91	0.78	0.81	0.97	0.59	0.56	0.75	0.69
		Legend	0.78	0.75	0.78	0.81	0.53	0.44	0.66	0.53
	Text	Labels	0.62	0.75	0.59	0.75	0.59	0.59	0.75	0.62
		Legend	0.72	0.62	0.72	0.69	0.59	0.5	0.69	0.5
Index-Middle	Default	Labels	0.66	0.84	0.75	0.84	0.72	0.66	0.75	0.81
		Legend	0.75	0.72	0.66	0.78	0.53	0.53	0.69	0.56
	Size	Labels	0.75	0.75	0.69	0.75				
		Legend	0.75	0.66	0.59	0.69				
	Brightness	Labels	0.78	0.88	0.72	0.75	0.66	0.62	0.84	0.72
		Legend	0.56	0.69	0.69	0.69	0.5	0.44	0.72	0.71
	Text	Labels	0.59	0.81	0.56	0.59	0.53	0.41	0.47	0.66
		Legend	0.62	0.66	0.69	0.78	0.31	0.44	0.72	0.69

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

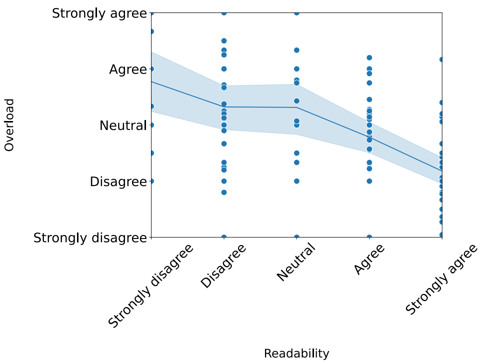
(b) **Appeal**

		Microgestures combo	Tap-Swipe				Tap-Hold			
		Spatial strategy	Overlapping		Juxtaposition		Overlapping		Juxtaposition	
		Family	A&B	MaS	A&B	MaS	A&B	MaS	A&B	MaS
Finger combo	Enhancement	Integration								
Index	Default	Labels	0.78	0.75	0.78	0.81	0.66	0.5	0.91	0.78
		Legend	0.81	0.84	0.75	0.81	0.75	0.56	0.84	0.81
	Size	Labels	0.78	0.78	0.78	0.66				
		Legend	0.81	0.72	0.81	0.84				
	Brightness	Labels	0.84	0.88	0.84	0.91	0.53	0.66	0.78	0.81
		Legend	0.78	0.84	0.78	0.88	0.44	0.5	0.78	0.75
	Text	Labels	0.84	0.75	0.84	0.81	0.72	0.88	0.94	0.94
		Legend	0.78	0.72	0.81	0.88	0.59	0.59	0.78	0.88
Index-Middle	Default	Labels	0.66	0.81	0.91	0.88	0.62	0.47	0.94	0.78
		Legend	0.69	0.59	0.72	0.91	0.5	0.5	0.91	0.72
	Size	Labels	0.72	0.84	0.69	0.78				
		Legend	0.69	0.66	0.78	0.78				
	Brightness	Labels	0.72	0.94	0.78	0.91	0.78	0.81	0.88	0.84
		Legend	0.72	0.84	0.78	0.88	0.5	0.56	0.78	0.75
	Text	Labels	0.72	0.88	0.72	0.88	0.59	0.72	0.84	0.81
		Legend	0.78	0.75	0.84	0.75	0.31	0.31	0.91	0.81

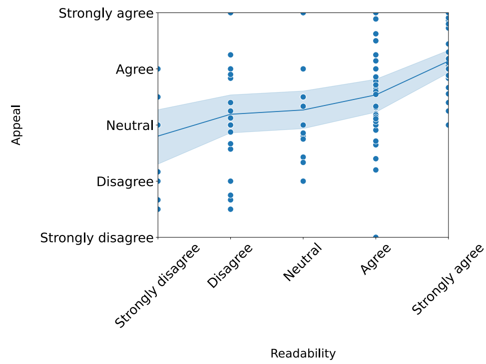
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

(c) **Readability**

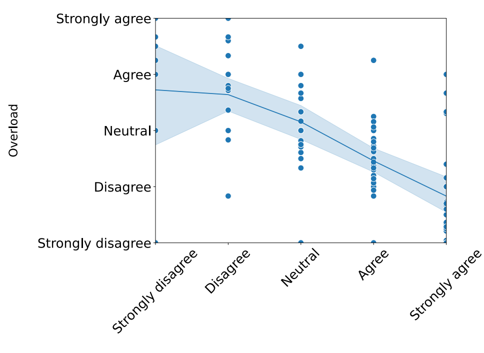
Table 3. Mean value of the easiness, appeal and readability measures for each representation among the 8 participants that have seen it



(a) Overload per readability



(b) Appeal per readability



(c) Overload per appeal.

Fig. 17. Correlation between measures : (a) overload per readability, (b) appeal per readability and (c) overload per appeal.

C User feedback on the help display (Experiment 2)

Experiment 2 used 4 measures to evaluate users impressions of a simultaneous representation of microgestures as an help display for an actual application. Figure 18 plots these measures while Figure 19 shows a screenshot of the experiment setup.

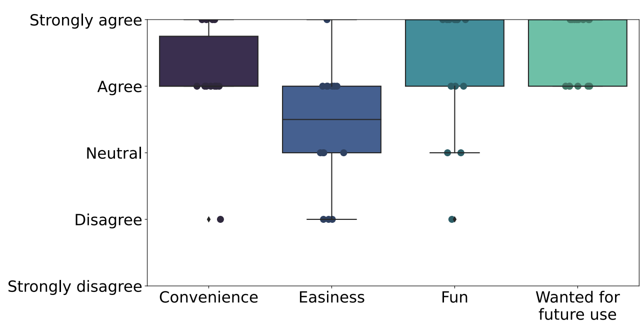


Fig. 18. User evaluations of convenience, easiness, fun and tendency to want this type of help for future microgesture-based applications.

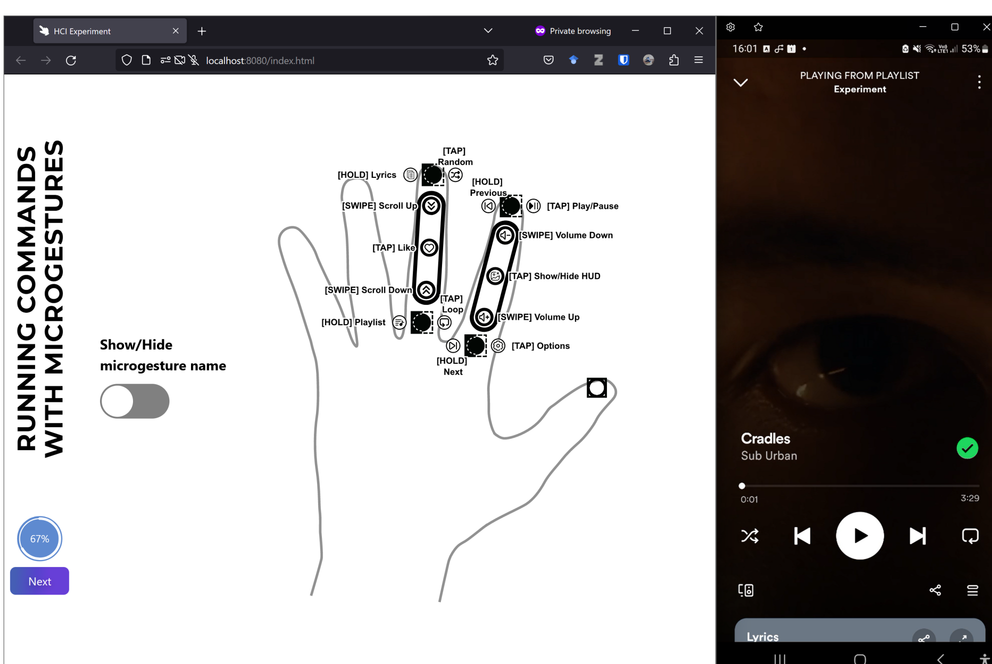


Fig. 19. Screenshot of the experiment with two distinct parts displayed.

D Differences between help displays in relation to the screen size (Experiment 3)

`Experiment 3` used 2 measures to evaluate each type of representation in relation to the screen size. Figure 20 focuses on the ease of finding the requested commands and Figure 21 focuses on the ease of understanding the microgesture related to the commands found.

Received February 2024; revised May 2024; accepted June 2024

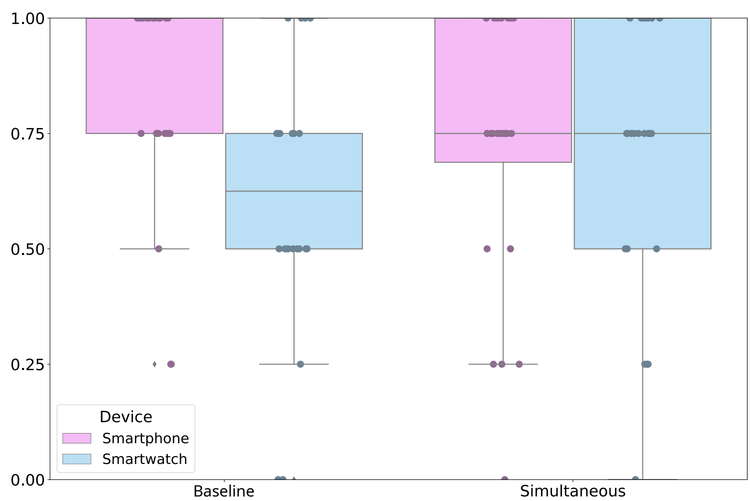


Fig. 20. User evaluations of ease of finding the requested commands in relation to the representation type and the screen size.

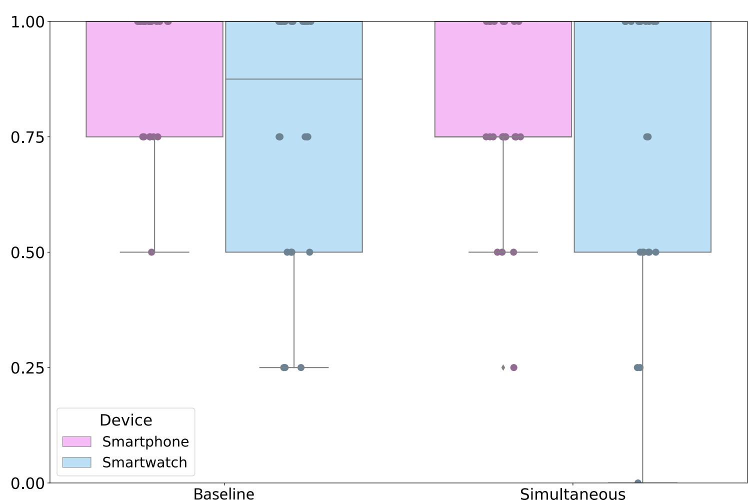


Fig. 21. User evaluations of ease of understanding the associated microgestures in relation to the representation type and the screen size.