Studying the Visual Representation of Microgestures

Vincent Lambert, Adrien Chaffangeon Caillet, Alix Goguey, Sylvain Malacria, Laurence Nigay

Studying the Visual Representation of Microgestures

VINCENT LAMBERT, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France
ADRIEN CHAFFANGEON CAILLET, 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

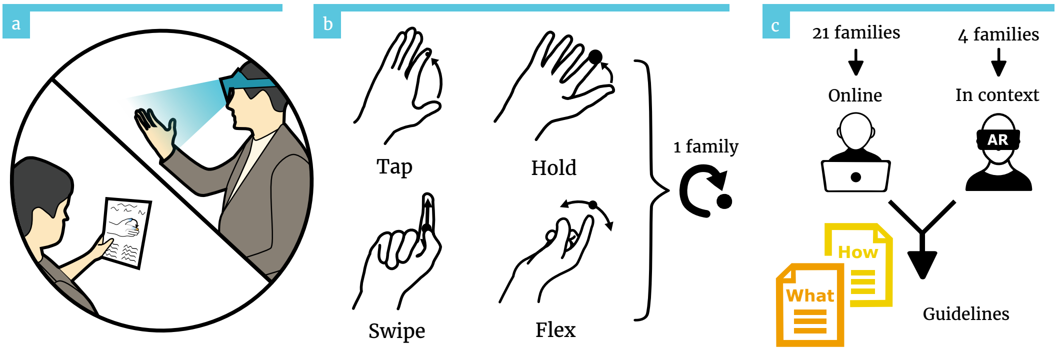


Fig. 1. Single-picture representations of microgestures. As part of microgesture learning strategies in an interactive system, such representations can be used in (a) various contexts, such as research papers and Augmented Reality applications. This study integrates (b) 4 microgestures, namely `tap`, `hold`, `swipe` and `flex` and proposes 21 families of representations sharing a common design among the microgestures. (c) These 21 families were tested in an online experiment, the 4 top ranked were further tested with an AR headset. Design guidelines for the representation of microgestures emerge from these two experiments.

The representations of microgestures are essentials for researchers presenting their results through academic papers and system designers proposing tutorials to novice users. However, those representations remain disparate and inconsistent. As a first attempt to investigate how to best graphically represent microgestures, we created 21 designs, each depicting static and dynamic versions of 4 commonly used microgestures (`tap`, `swipe`, `flex` and `hold`). We first studied these designs in a quantitative online experiment with 45 participants. We then conducted a qualitative laboratory experiment in Augmented Reality with 16 participants. Based on the results, we provide design guidelines on which elements of a microgesture should be represented and how. In particular, it is recommended to represent the actuator and the trajectory of a microgesture. Also, although preferred by users, dynamic representations are not considered better than their static counterparts for depicting a microgesture and do not necessarily result in a better user recognition.

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

Additional Key Words and Phrases: Microgesture; Microgesture representations; AR; Discoverability

Authors' addresses: Vincent Lambert, vincent.lambert@univ-grenoble-alpes.fr, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France; Adrien Chaffangeon Caillet, adrien.chaffangeon@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.

Authors' version
https://doi.org/10.1145/3604272

ACM Reference Format:
Vincent Lambert, Adrien Chaffangeon Caillet, Alix Goguey, Sylvain Malacria, and Laurence Nigay. 2023. Studying the Visual Representation of Microgestures. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 225 (September 2023), 35 pages. https://doi.org/10.1145/3604272

1 INTRODUCTION

Single-hand microgestures are a subset of gestures that can be performed by subtly moving one or more fingers while moving neither the arm nor the wrist. They bring together sought-after characteristics such as versatility in the context of use (eye-free, hand-free or not), ease of execution (avoidance of the "Gorilla arm" effect [31]), and a better social acceptance than other gesture-based interactions (e.g. whole-hand, arm, body gestures) [13, 15, 40].

The domain of microgesture interaction is very dynamic with a focus made on elaborating relevant microgesture sets through elicitation studies [13, 40, 45] and evaluating the end-users' performances [9, 34, 55, 66]. Another area of research has been the design and implementation of systems and devices that sense and recognize such microgestures [30, 51, 68]. However, as for any other gesture-based interaction, using microgestures as inputs requires to be aware of which gestures are available and which commands they may activate. To this extent, researchers and practitioners face the same need for common guidelines on how to intelligibly *present* the microgestures they use.

As shown in Figure 2a, microgestures are typically represented with multiple images that depict the movement step-by-step. Some visual cues such as arrows or colored elements can be used to show the movement done from one step to another. However in most cases, as presented in Figure 2b, these cues are used with a single image that can then represent the initial, intermediate or end position of the hand. This diversity of representations can be confusing, especially when it comes to distinguishing between two similar microgestures. Practitioners encounter similar issues as they tend to use crib-sheets along with video tutorials to depict gesture-based interaction [3, 46].

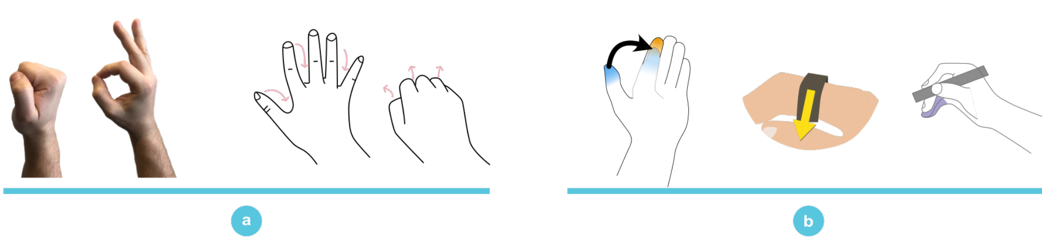


Fig. 2. Depictions of microgestures in research papers.
a) Multi-step microgesture depictions inspired from [24, 40].
b) Single-step microgesture depictions inspired from [9, 10, 56].

McAweeney *et al.* [48] proposed design observations and a taxonomy of design elements to statically represent any kind of gestures, including few microgestures. They asked groups of participants to design gesture representations with multiple iterations. They analyzed the final representations and extracted principles from their observations. As a use case, the authors redesigned the representation used by Microsoft for the `tap` microgesture to reduce it to a single image as it "avoids redundant and excessive motion expression". Even though they did not provide precise guidelines directly usable for microgesture representation, their results associated to Antoine *et al.* taxonomy of static illustrations [2] give directions on the representation of gestural interaction in general. For their part, Mackamul *et al.* [44] investigated how different button designs could communicate in-place touch inputs, but these inputs are radically different from microgestures as they are stationary and performed on a touchscreen. To our knowledge, no other study has focused on how to represent a microgesture nor produced guidelines on how a microgesture should be represented.

Therefore, we compare distinct visual cues which constitute the basis of microgesture representation for both static (e.g. in a crib-sheet) and dynamic image (e.g. in video tutorials) and aim to distinguish between what is false common sense and what is supported by scientific evidences.

We chose to focus on single-image representations all using the same initial hand posture, i.e. an opened hand. Our contribution is the creation of 21 families of representations and their design rationale. Each family represents 4 microgestures with similar visual designs (Figure 1b). We distinguished the *static* (i.e not animated) and the *dynamic* (i.e animated) states. Each family is thus composed of 4 *static* and 4 *dynamic* representations sharing consistent designs.

These families have been assessed by 45 participants in an online experiment. Then, the best 4 families have then been tested in context using an Augmented Reality (AR) headset with the ambition of refining our quantitative results by qualitative information. The results allow us to establish design guidelines on the representation of microgestures (Figure 1c). Our findings focus on which part of the microgesture should be described and how they should be represented. We confirmed that the reasonable hypothesis of showing a maximum of information to create a good representation applies to microgestures. We also nuance this result by showing representations that suggest the movement without explicitly showing its trajectory. Our results tend to support arrow-based representations as a good practice. We also propose a high-level discussion and guidelines helping researchers and practitioners to better choose visual cues and their characteristics, e.g. size, color, shape.

Our contributions are four-folds:

- 21 families regrouping the representations of 4 microgestures and their design rationale
- Results of an online user study comparing the 21 families both statically and dynamically
- Results of a laboratory user study testing the 4 best families in an AR context
- A set of guidelines for the design of microgesture representations

2 CONSIDERING COMMON MICROGESTURES

Our work studies the representation of microgestures with a single image. In this section, we first review the variety of microgestures and their characteristics. We then define a set of the most commonly used and studied microgestures.

2.1 Microgestures in the literature

Hand microgestures, simplified as microgestures in this paper, are described as fast and subtle gestures [13, 56] performed by the hand on itself [13, 59] or on an object [56, 69].

From the Sayre Glove [18], which could detect finger flexion, to FingerInput [59], a gesture recognition system using depth sensing and a convolutional neural network, many sensing technologies have been developped [16, 30, 67]. Based on elicitation studies [13, 40, 45, 56], the microgesture diversity has grown exponentially and has become substantial. Chaffangeon Caillet *et al.* enumerate more than 118 different microgestures and propose μGlyph [10, 11], an expert notation to describe them according to their movement(s) and execution context(s). In order to structure our exploration of the existing microgestures, we chose to categorize them using the two contexts introduced by μGlyph : in contact with a surface, represented by •, and in the air, represented by ◦. This leads to the existence of the following three types of microgestures:

- **Context-switching microgestures** (○● or ●○) which begin in a given context and end in another, e.g. a finger in the air coming into contact with the palm of the hand;
- **In-context microgestures** (●● or ○○) which keep the same context throughout, e.g. a finger draging onto another finger;
- **Stationary microgestures** (● or ○) which keep the same context and the same position, e.g. a finger increasing pressure onto another finger or simply a still finger.

Context-switching microgestures have diverse variations but usually begin with a `touch`, i.e. ○●, and/or finish with a `release`, i.e. ●○, as shown by the first two tables of μGlyph appendix [10]. `Tap`, i.e. ○● then ●○, is the most common context-switching microgestures [13, 22, 40, 49, 51, 55, 56, 59, 70].

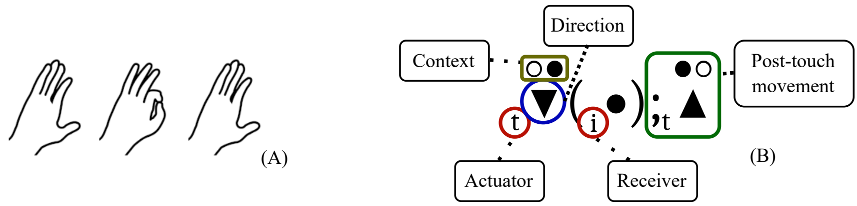
In-context microgestures are characterized by their movement which is either a line or a complex shape. `Swipe`, `slide` and `drag`, i.e. ●● linear movements, are the most common in-context microgestures [13, 22, 40, 49, 51, 55, 56, 59, 70]. These microgestures often work in pairs by considering one direction and its opposite. Thus, they are often distinguished into two categories: horizontal `swipe` [9, 59], when performed sideways across one or multiple fingers, and vertical `swipe` [13, 19, 32, 59], when performed up-/down-wards along the face or the side of a finger. `Stretch` and `flex` are equivalent to the vertical `swipe` performed in the context ○○. They are not as common as their ●● counterpart, but are used in some studies [38, 59]. In this paper, we use the term `flex` to refer to both `stretch` and `flex` as it is the most commonly used term [38, 52, 59]. Finally, microgestures made of more complex and various shapes, e.g. a circle or a square, drawn either in contact or in the air are called `draws`. They form a minority of the microgestures favored by end-users who prefer simpler movements as `taps` or `swipes` in elicitation studies [13, 45, 56].

Stationary microgestures are microgestures where the finger do not move from its original position. As specified in μGlyph [10, 11], it is still possible to perform subtle movements while being in the context ●. For instance, `press` is an augmentation of the pressure on a surface by subtly increasing the force applied by a finger on this surface. `Press` is the most common stationary microgesture [10, 11, 56, 68, 69]. μGlyph also allows to describe dwell time by representing a static finger for a specified duration.

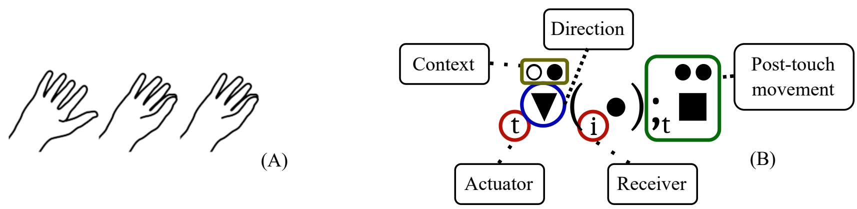
Finally, those types of microgestures can be combined to create more complex microgestures. In particular, dwell time is always part of a more complex microgesture such as `hold`, i.e. a `touch` followed by a dwell time.

2.2 Chosen microgestures

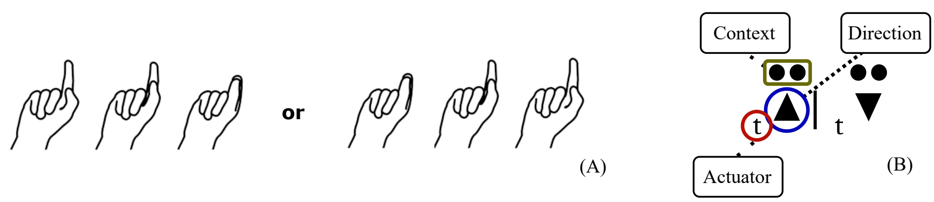
To embrace the diversity of microgestures, we have chosen microgestures covering the different contexts presented above. For context-changing microgestures, we decided to keep the omnipresent `tap`, which covers both ○● and ●○ contexts. For in-context microgestures, we kept both `swipe` and `flex` to cover both ●● and ○○ contexts. We chose to use their vertical version ▲| ▼ as they are rarely executed side-ways ◄| ► [13, 32, 59]. We did not consider the `draw` microgesture as it can be modelled by a sequence of `swipes` and `flexes`. Therefore, we first need to establish satisfying representations and guidelines for `swipes` and `flexes` before tackling more complex gestures such as `draws`. For stationary microgestures, we decided to keep the `hold` microgesture to cover the context ● instead of the more present `press`. This was done for two reasons. First, the `hold` is similar to the `tap`, as they both start with a `touch`, which creates an ambiguity for creating coherent yet different representations for both microgestures. Second, the `hold` also allows us to cover microgestures composed of two different elementary microgestures, i.e. a context-changing microgesture and a stationary microgesture as illustrated by Figure 3b. Finally, we do not cover the context ○ as no existing microgesture uses this context [10, 11].



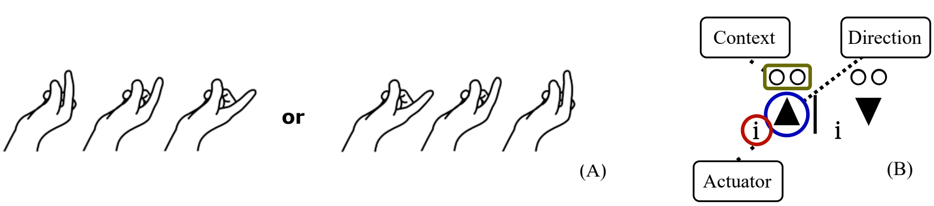
(a) `Tap` ${}_{t}\overset{\circ\bullet}{\blacktriangledown}({}_{i}\bullet);{}_{t}\overset{\bullet\circ}{\blacktriangle}$



(b) `Hold` ${}_{t}\overset{\circ\bullet}{\blacktriangledown}({}_{i}\bullet);{}_{t}\overset{\bullet}{\blacksquare}$



(c) `Swipe` ${}_{t}\overset{\bullet\bullet}{\blacktriangle}|{}_{t}\overset{\bullet\bullet}{\blacktriangledown}$



(d) `Flex` ${}_{i}\overset{\circ\circ}{\blacktriangle}|{}_{i}\overset{\circ\circ}{\blacktriangledown}$

Fig. 3. (A) The 4 considered microgestures with their (B) µGlyph description.

With µGlyph, every microgesture can be described with combinations of movement and context glyphs. The movement is represented by a triangle pointed in the direction of the movement, e.g. ▼. The absence of movement is represented by a square, e.g. ■. The initial and final contexts of the movement, e.g. ○●, are positioned above the movement glyph. A sequence of movements is separated by a " ; " and a choice between two movements is represented by "| ". Therefore, a finger going onto another one then releasing a contact, in other words a `tap`, is written as $\overset{\circ\bullet}{\blacktriangledown};\overset{\bullet\circ}{\blacktriangle}$. For any microgesture, µGlyph also allows us to specify which fingers act as *actuator*, i.e. the finger initiating the microgesture, and *receiver*, i.e. the finger involved in the movement initiated by the actuator. For our study, we decided to use the thumb, represented by the symbol "t ", and the index finger, represented by the symbol "i ", as actuators and receivers. Indeed, there is an overwhelming presence of thumb-to-finger microgestures [9, 13, 19, 23, 30, 32, 35, 41, 51, 59, 68]. While the thumb can `tap`, `swipe` and `hold` on a finger or a specific part of the finger such as the bottom phalanx, according to [19] and Digitspace, the most comfortable area is the index fingertip. As very few studies [52, 67] differentiate the `flex` based on the moving joint(s) of the finger, we chose to keep the index finger as a whole as the actuator for `flexes`.

Figures 3a, 3b, 3c and 3d represent the resulting set of microgestures that we consider in our study. Each microgesture is represented by a step by step representation and the associated μGlyph description.

3 VISUALLY REPRESENTING MICROGESTURES

On the one hand, μGlyph [10, 11] is a symbolic expert notation that allows a complete description of the motion characteristics. On the other hand, the illustrations of microgestures in papers typically depict a more realistic hand drawing. μGlyph representations have been designed for researchers and need a training to be usable [11]. Furthermore, μGlyph does not provide a visual representation of the actuator trajectory whereas it is often represented in the illustrations of the papers. For instance, DigitSpace [32] uses a combination of straight arrows and circles for the `swipe` microgesture whereas PinchWatch [43] uses curved arrows.

We based our review of microgesture representations used in the literature on two design frameworks: 1) the set of design elements to present the motion of a gesture proposed by McAweeney *et al.* [48]; 2) the taxonomy of interaction illustrations defined by Antoine *et al.* that focuses on static illustrations [2]. Our review of existing microgesture representations led us to define a categorization of visual cues presented in Table 1.

The design space resulting from this categorization is too large to be systematically explored. We chose to cover it with representations using visual cues extracted from Table 1. We enriched our approach with inspirations from various fields beyond the HCI field. The resulting representations have been grouped into distinct design "families". Eventually, we deepened our research work on Bertin's variables introduced in *Semiology of Graphics* [8] as a mean to take a step back from the designed families in order to finalize them.

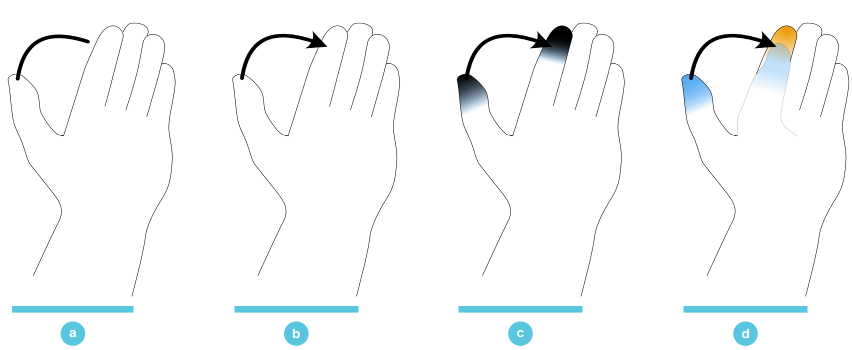


Fig. 4. Decomposition of a static illustration into visual cues (example reproduced from [10]). (a) Trajectory with a line. (b) Direction with arrow tip. (c) Actuator & Receiver with black organic shapes. (d) Emphasis tools with colors and opacity.

Trajectory	Lines	Simple line	[5, 9, 10, 13, 15, 21, 22, 49, 50, 55, 56, 59, 64, 66, 69]
		Curve line	[5, 10, 13–15, 21, 24, 32, 40, 49, 50, 55, 56, 59, 64, 68, 69]
		Broken line	[10, 14, 15, 21, 32, 49, 55, 56, 59]
	Stroboscopic effect		[5, 10, 15, 49, 64, 68]
Direction	Arrows	Simple-headed	[5, 9, 10, 13, 15, 21, 24, 32, 40, 49, 50, 55, 59, 64, 66, 68, 69]
		Double-headed	[5, 13, 22, 40, 49, 50, 56, 59, 64, 69]
	Motion lines		[15, 55]
Actuator & Receiver	Shapes	Organic	[15, 49, 55, 56, 59, 64, 66, 68]
		Geometric	[9, 15, 22, 32, 40, 59, 64, 66]
	Icon		[5, 14, 56, 64, 68]
Emphasis tools	Stroke	Dash pattern	[32, 49, 55, 59, 64, 66, 69]
		Color	[9, 14, 15, 21, 24, 32, 40, 50, 55, 59, 64, 66, 68]
	Fill	Opacity	[5, 10, 15, 49, 68]
		Color	[9, 10, 15, 22, 32, 50, 56, 59, 64, 66, 68]
	Text		[32, 40, 49, 64, 68]

Table 1. Categorization of the visual cues used in the literature

3.1 Representations used in HCI research papers

McAweeney *et al.* conducted a partnered elicitation study on the most commonly used graphical elements to represent different parts of a gesture interaction including *motion*, i.e. the movement performed by an actuator, and *touch*, i.e. the contact of an actuator and a receiver [48]. The results of their study suggest that the most commonly used graphical elements are 1) "ghosts", i.e. stroboscopic images of the fingers, or arrows for *motion* and 2) "touchpoints" or colors for *touch*. Based on the literature, including the observations of McAweeney *et al.*, Antoine *et al.* proposes a taxonomy of interaction illustrations [2]. The taxonomy is built on categories of visual cues including "Lines", "Arrows", "Contact shapes" and "Effects". These 4 categories can be used to represent different parts of a microgesture. The *motion* of a microgesture is composed of a **trajectory** often depicted by Lines and a **direction** indicated mainly by Arrows. The *touch* event of a microgesture is related to both the **actuator** and the **receiver** depicted by Contact shapes. Other **emphasis tools** are also used to further represent the *motion* or the *touch*, e.g. parts of a microgesture represented with Effects elements. Figure 4 explains how these 4 categories make up a representation of a microgesture.

Trajectory is mainly represented by different kinds of Lines. Curves and broken lines are used mainly to represent `draw` microgestures, e.g. circle or square `draws`, and also movements of the tip of a finger [13, 50, 55, 56, 59]. Another strategy to depict the trajectory in *static* images, is the use of a *stroboscopic* effect to display multiple positions through time [17, 48]. McAweeney *et al.* study indicates that this strategy has been selected by 25.3% of their user groups to represent the *motion*.

Direction is mainly indicated by Arrows. Associated with Lines or more rarely used as a pattern to shape a chevron [69], arrows indicate the direction of a microgesture as shown in Figure 4b. Motion lines[1], which depict motions in comics, can also be used in papers to convey the direction of the movement [15, 55].

Actuator and receiver are often shown through Contact shapes. These shapes can vary along a continuum from shapes that exactly match the outline of the finger or hand part described, i.e.

[1]See *Motion lines* [Wikipedia]

organic shapes [55, 56, 59, 64], to generic geometric shapes above or next to the part of the finger or hand described [32, 51, 59]. Geometric shapes can also be replaced by icons or pictograms allowing for a trade-off between visual load and depiction of additional characteristics of the considered movement [54].

Emphasis tools are Effects that modify the aspect of a visual cue. Examples of aspects of a visual cue include the line type, the filling pattern, the color, and the level of transparency [2, 13, 45, 56, 59, 69]. Practitioners vary the line type, the filling pattern, the color, the level of transparency and other characteristics of the visual cues they use [2, 13, 45, 56, 59, 69]. Antoine *et al.* categorized emphasis tools under their Effects and Emphasize sections. As they are applied to Lines, Arrows and Contact shapes, emphasis tools can be used to reinforce the mental association between two elements. This has been done for context-switching microgestures such as `taps` or `holds`. Soliman *et al.* chose a color association between the actuator and the receiver [59] whereas Chaffangeon Caillet *et al.* preferred a shape association [10, 11]. This allowed them to represent a microgesture without specifying the trajectory.

Table 1 summarizes the above categorization of the visual cues used in the literature to depict microgestures.

3.2 Representations from other fields

We observed during our review process that some representations used visual cues existing in other fields in which movements are commonly represented. This widens the already large design space depicted in Table 1. For example, the field of *origami* has been used by researchers of many fields as a source of inspiration [33] and should serve as an inspiration for us as *origami* implies finger movements to fold papers. The work of Akira Yoshizawa[2], considered as a master of *origami* art, is a central reference. It proposes visual cues now used by every origami book as the broken arrow for the "fold then unfold" step [1]. *Video-games* define another field which uses various cues to indicate points of interest through their HUDs [62]. In *sports* and early *bio-mechanical* studies, the use of chronophotography (or stroboscopic image) was omnipresent to describe body motions [53, 63]. Other research fields such as *mechanics* and *electromagnetism* also rely on sketches to describe the motion of metal pieces or electrons [60]. On another note, we also looked at road signs which are globally standardized[3] and therefore should be little impacted by cultural differences.

3.3 Chosen microgesture representations

The design space of possible cues to depict microgesture is too large to be systematically explored. We cannot build an experimental protocol to test every chevron shape or icon type. Consequently, we decided to cover this large design space with a limited amount of representations. By doing so, we aim at providing first insights on what design properties work or not. In order to evaluate many visual cues for the 4 chosen microgestures, we group representations into families. Each family is a set of representations created to represent the 4 microgestures (`tap`, `swipe`, `flex` and `hold`).

The creation process of the families took inspiration from HCI research papers and the other inspiration fields as described above. We aimed to test different subsets of the categories introduced by Table 1, chosen according to their relevance. The key principle is that a visual cue widely used in the literature , e.g. an arrow, should appear in multiple families to study the influence of the other categories, e.g. the line or the color, on its relative importance. Consequently, we created 21 families:

[2]See *Akira Yoshizawa* [Wikipedia]
[3]See *Vienna Convention on Road Signs and Signals* [Wikipedia]

- 3 families are from HCI papers and use designs distinct from our inspiration fields: $_{A\&B}$ ([26, 37] - arrow and ball), $_{DpA}$ ([22, 56, 59] - double path arrow) and $_{AtC}$ ([45, 58] - arrow tip with circle).
- 11 families are inspired from these same papers but have been designed with a direct correspondence to concepts from the above mentioned inspiration fields. For example the $_{MaS}$ family uses the visual association of a disk and a circle already used by Chaffangeon Caillet *et al.* [10, 11]. The authors use this association to communicate a change of context provoked by the movement of a finger but this association can also represent the finger itself. The metaphor of a disk going into a circle is also used in early shape sorter games. The other families of this group are the following ones : $_{BA}$ ([10, 55] - origami broken arrow), $_{At}$ ([39] - origami pressure arrow), $_{Chv}$ ([69] - direction road sign), $_{Chp}$ ([15, 34] - sport chronophotography), $_{HiZ}$ ([23, 55, 59] - video-games highlighting), $_{HiB}$ ([56, 59] - video-games highlighting), $_{DsL}$ ([66] - origami lines and dashes), $_{CwS}$ ([64] - computer and smartphone cursors), $_{MSt}$ ([59] - movement in arts, notably comics), $_{EmL}$ ([15, 55] - movement in arts, notably comics).
- 7 families are inspired by other fields than HCI and are therefore new, i.e. have no link with existing representations used in HCI papers: $_{DeA}$ (origami and mechanics arrows), $_{DsB}$ (art dotted paths), $_{CnS}$ (computer and smartphone cursors), $_{Ri}$ (video-games rings), $_{ElC}$ (electric charges), $_{Gra}$ (sports and video-games progress meters) and $_{Tar}$ (darts).

Annex A details how these 21 families cover the categories presented by Table 1.

We also had to consider the use of animated representations [3, 46]. The words *static* and *dynamic* are currently used in the literature [6, 12, 25, 29] to distinguish still cues and their animated counterparts. Gao [27], Yeung *et al.* [71] and Mayer *et al.* encourage to “animate [whenever] possible” [47]. We therefore created a *dynamic* version of the 21 families. Since the number of possible outcomes is also virtually infinite, we followed this simple rule: the “dynamization” process should only impact a subset of Bertin’s variables – namely position, value and color as Sharma *et al.* did with the one microgesture they “dynamized” in their supplementary material [59].

In summary, with respect to the different visual cues in the HCI literature and other thematic fields, we narrowed down the resulting large design space by considering 21 families in this study. Each family is composed of 4 microgesture representations (one for each `tap`, `swipe`, `flex` and `hold`) derived for 2 states (*static* and *dynamic*). Thus each family contains 8 representations sharing a consistent design.

Table 2 presents the 21 families, showing for each family the static representation of the 4 microgestures. For the sake of conciseness and clarity, the table has been designed to serve as a one-page guide that could help in the exploration of the following sections of this paper. To quickly identify the families in the text or figures, we chose to identify each family by a symbol, a name and an abbreviation. Symbols are similar to icons and like them must be "compact" [6]. Thus, we chose to focus on the visual cues used in the `tap` representations. All the resources and instructions needed to reuse these families are available as supplementary materials.

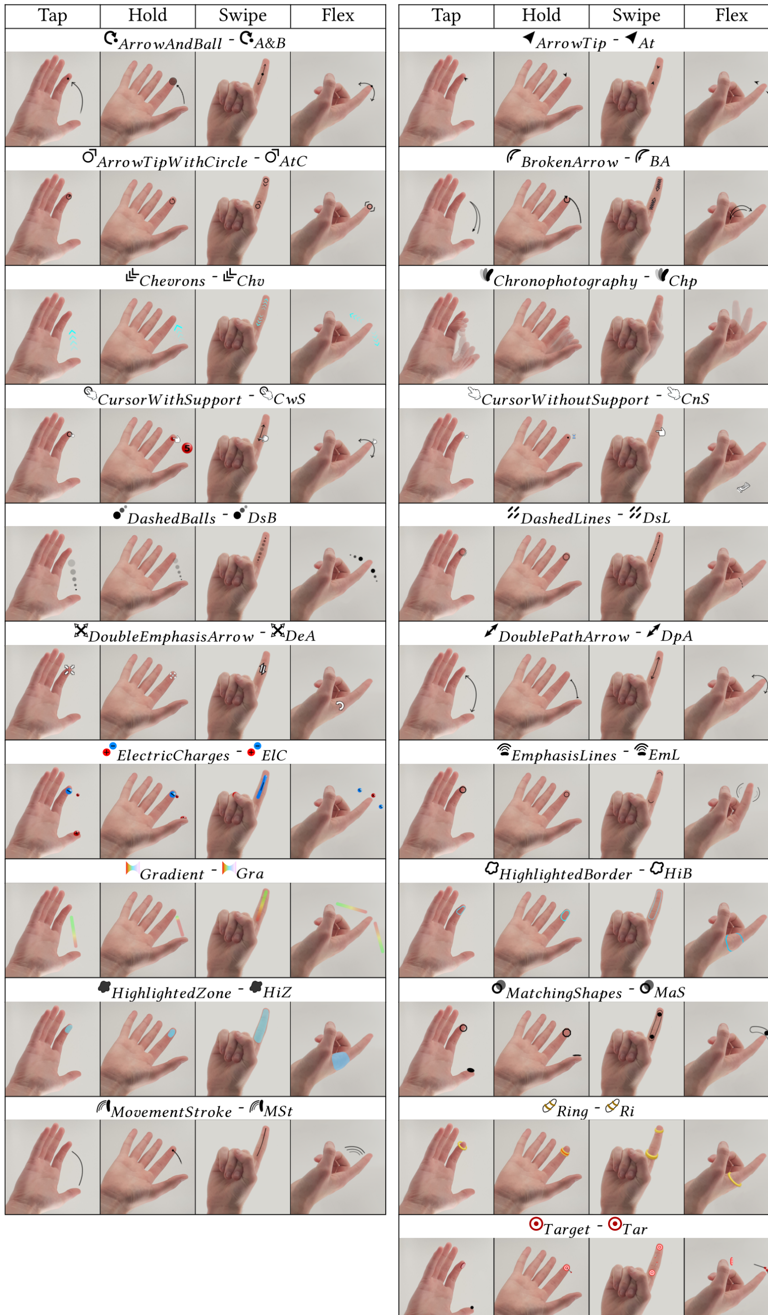


Table 2. Static representations of the 21 families for the `tap`, `hold`, `swipe` and `flex` microgestures.

4 EXPERIMENT 1 : COMPARING THE REPRESENTATIONS OF MICROGESTURES

In this work, we aim to provide initial insights on how to efficiently visually represent microgestures. As stated earlier, such representations are essential in many contexts from researchers presenting their results in academic papers to system designers developing tutorials for novice users in an Augmented Reality context. As a first step, we conducted a study evaluating the 21 families of representations described in the previous section when displayed on a conventional computer screen. We deployed an online form in which respondents were asked to score how efficiently each graphical representation conveys the associated microgesture.

4.1 Participants

We recruited 45 unpaid volunteers whom age ranged from 13 to 76 years old (mean=32, std=14). We recruited them through Facebook posts on student and city groups, and messages sent to mailing lists. 52% were women, 43% were men, one person identified as non-binary. We asked them to score their familiarity with AR (6-point Likert scale), microgestures (5-point Likert scale) and our inspiration fields (binary variables).

4.2 Apparatus

We did not impose specific platforms on which to fill up the form. Participants could use their smartphone or their computer. The online form is available as a supplementary material.

4.3 Procedure

(1) Participants were first introduced with the terms of the topic, i.e. microgesture, representations and Augmented Reality. They then scored their familiarity with both microgestures and AR. Before starting the evaluation phase, we instructed them to imagine the representations as if they were in an AR context.
(2) Participants were asked to evaluate the ability of each representation to portray a given microgesture. These evaluations were performed using a 6-point Likert scale[4]: Very bad, Bad, Quite bad, Quite good, Good, Very good. Participants could optionally add a comment to explain their scores for all representations. In total, they scored 4 microgestures × 2 states × 21 families = 168 representations. We organized this phase in 8 scoring pages, each corresponding to a microgesture (i.e. `tap`, `swipe`, `flex`, `hold`) and state (i.e. *static*, *dynamic*) combination. At the beginning of each of those pages, an animated GIF and a textual description introduced the microgesture to evaluate in the page.

[4]There are no confirmed differences between even and odd numbered scales [57]. We used a 6-point Likert scale to force participants to decide whether they felt positive or negative about the representations.

Fig. 5. Likert scale used with the associated representation of a microgesture.

(3) Finally, participants answered a small demographic survey and indicated with check-boxes their familiarity with the thematic domains used as inspiration, e.g origami and video-games. They also had the opportunity to write down any final comments on the study.

To observe if the presentation order of states and microgestures has an impact on scores, we applied a controlled randomization of the representations order. For each participant, we randomly selected which state to start the study with, resulting in two categories *static 1st* or *dynamic 1st*. We then randomly generated a microgesture order. Participants thus saw all 4 pages corresponding to their first state following the generated microgesture order, and then saw all 4 pages corresponding to the second state following the same microgesture order. In each of these pages, the order of appearance of the families was randomized. For instance, the 8 scoring pages could have been organized as followed : (*static*, `tap`)-(*static*-`flex`)-(*static*, `hold`)-(*static*, `swipe`)-(*dynamic*, `tap`)-(*dynamic*-`flex`)-(*dynamic*, `hold`)-(*dynamic*, `swipe`).

4.4 Quantifying the information provided for a microgesture

We mentioned earlier that the representations used in the literature do not explicitly convey all the information about the movement of the microgesture, e.g. finger "ghosts" used by McAweeney *et al.* show the trajectory of the finger but not the direction of the movement. This raises the following question: Does the amount of information represented influence the appreciation of a family? To quantify the amount of information represented, we introduce a new metric: the *explicitness*.

µGlyph decomposes microgestures according to different features, i.e. direction, execution context, actuator, receiver and additional movement characteristics. Besides, the literature usually represents the trajectory of the microgesture. Therefore, we define the *explicitness* as the ratio of the number of represented features over the total number of features, for a given microgesture. Figure 6

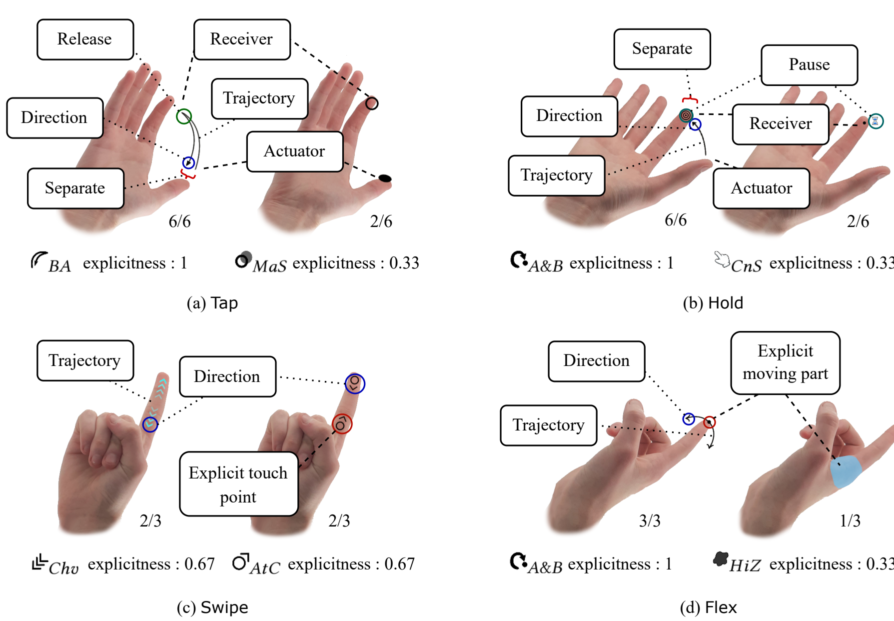


Fig. 6. Explicitness decomposition: (a) `tap`; (b) `hold`; (c) `swipe`; (d) `flex`.

summarizes the features used to compute the explicitness ratios for the 4 microgestures. The explicitness ranges from 0, i.e nothing is shown, to 1, i.e. everything is depicted. As visible on Figure 6a the `tap` $_{BrokenArrow}$ representation has an explicitness of 1. The broken arrow explicitly shows the trajectory of the movement (1/6) and implicitly specifies the actuator (2/6) and the receiver (3/6) with its ends. Furthermore, the direction of the movement is indicated by an arrow (4/6) and the release information is given (5/6) with a separate, i.e. not superimposed, element being here the broken part of the arrow (6/6). On the contrary the $_{MatchingShapes}$ family only shows the actuator (1/6) and the receiver (2/6). Thus, its explicitness value is 0.33. Annex C (Tables 6, 7, 8 and 9) details the calculation of explicitness for each microgesture representation, classified by family.

4.5 Results analysis

Following the recommendations for "fair statistical communication in HCI" [20], we use 95% confident intervals computed using bootstrapping, with σ the standard deviation and N the size of the considered sample. Using the TIP 24 of the guidelines [20], if the CIs visually overlap by more than 1/4 of their length, then we can use a similar reasoning to that of p-value less than 0.05. Moreover, when adequate, we report the standardized effect size using Cohen's d parameters. Bootstrapped CIs and Cohen's d were computed with the `scipy.stats` package for Python with the default parameters.

To perform our analysis, our Likert scale with *Very bad* to *Very good* was encoded using numbers from 1 to 6.

We did not observe any impact of the microgesture order or the state order (*static 1st* or *dynamic 1st*) on the Likert scale score. For the sake of clarity and conciseness, the plots to validate this statement have been moved to the Annex B.

In addition, we initially intended to assess whether an individual's familiarity with a certain domain could have an impact on the scores of representations inspired by that domain or representations in general. However, we did not observe any significant impact of the self-rated familiarity variables on our results. Consequently, in the following we consider participants altogether regardless of their familiarity or expertise in microgestures, AR and the thematic domains used as inspiration.

4.5.1 Families preference. In general, **most families are not satisfying enough.** Indeed, 9/21 obtain strictly negative scores, i.e. confidence intervals below the *Quite bad* score, see Figure 7a. On the contrary, only the $_{Chronophotography}$ family has strictly positive scores, i.e. confidence intervals above the *Quite good* score.

We observe a clear difference between the arrow-based families showing the trajectory of the microgesture ($_{BrokenArrow}$, $_{DoublePathArrow}$, $_{ArrowAndBall}$ and $_{Chevrons}$) and the ones not showing any trajectory($_{ArrowTip}$ $_{DoubleEmphasisArrow}$, and $_{ArrowTipWithCircle}$), i.e. their CIs do not overlap. Therefore, it seems that **arrow-based families are preferred as long as they show the trajectory**.

The overall score of the families does not differ between microgestures: `tap` (Mdn = 3.00, CI = [2.88, 3.19], M = 3.00, SD = 0.72, N = 45), `hold` (Mdn = 2.95, CI = [2.81, 3.19], M = 3.02, SD = 0.77, N = 45), `swipe` (Mdn = 2.98, CI = [2.69, 3.17], M = 2.83, SD = 0.77, N = 45) and `flex` (Mdn = 3.17, CI = [2.93, 3.36], M = 3.06, SD = 0.70, N = 45). However, while **the families ordering for the 4 types of microgestures are similar**, some particularities emerge, see Figure 7(d-g). For instance, we observe that the $_{ArrowTip}$ family obtain significantly better results for the `flex` (Mdn = 4.00, CI = [3.50, 4.00], M = 3.73, SD = 0.94, N = 45) than for the other microgestures (Mdn = 2.00, CI = [2.00, 2.50], M = 2.41, SD = 1.06, SD = 1.06, N = 45 for the `tap`, Mdn = 2.00, CI = [2.00, 2.00], M = 2.13, SD = 0.97, SD = 0.97, N = 45 for the `hold` and Mdn = 2.50, CI = [2.50, 3.00], M = 2.66, SD = 1.1, SD = 1.1, N = 45

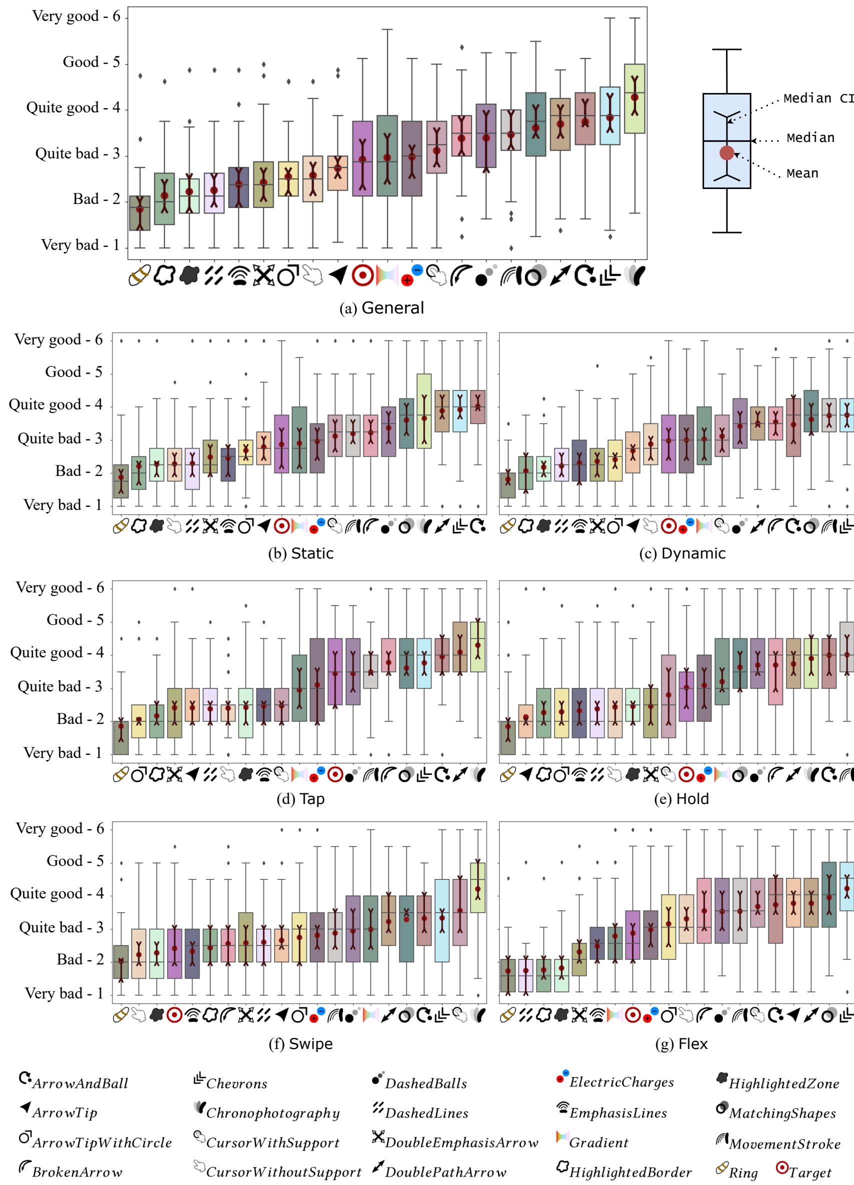


Fig. 7. Reported quality scores of the representations with their respective quartiles, mean (red dot) and 95% bootstrapped CI for the median per family: (a) in general; (b) for *static* state only; (c) for *dynamic* state only ; (d) for the `tap` only; (e) for the `hold` only; (f) for the `swipe` only; (g) for the `flex` only.

for the `swipe`) with a large effect size (d>1). The same goes for the *CursorWithSupport* family for which the results obtained for the `swipes` (Mdn = 3.50, CI = [3.00, 4.50], M = 3.56, SD = 1.4, N = 45) and `flexes` (Mdn = 3.50, CI = [3.50, 4.00], M = 3.63, SD = 1.19, N = 45) show a medium effect size (d>0.5) when compared to those obtained for `taps` (Mdn = 2.50, CI = [2.00, 2.50], M = 2.47, SD = 1.01, N = 45) and `holds` (Mdn = 2.50, CI = [2.00, 3.50], M = 2.8, SD = 1.37, N = 45).

4.5.2 Static vs dynamic. In general, ***dynamic* representations (Mdn = 3.12, CI = [2.94, 3.30], M = 3.00, SD = 0.64, N = 45) are not preferred over *static* representations (Mdn = 3.08, CI = [2.85, 3.15], M = 2.95, SD = 0.78, N = 45).**

However, for the specific case of the $_{Chronophotography}$ family, participants had extreme and opposite appreciations. In fact, its *dynamic* version is by far the most preferred family (Mdn = 5.00, CI = [4.75, 5.50], M = 4.90, SD = 1.15, N = 45) as depicted by Figure 7c, well beyond the second family ($_{Chevrons}$) (Mdn = 3.75, CI = [3.50, 4.00], M = 3.76, SD = 0.97, N = 45). However, its *static* version received much lower results (Mdn = 3.75, CI = [3.00, 4.25], M = 3.66, SD = 1.43, N = 45) as visible on Figure 7b. Using transparent fingers thus appear to be a good idea to depict a microgesture as long as they are *dynamic*. $_{Chp}$ *static* representations having "too many superimposed images" (P25), it makes this family both ambiguous and confusing for the `tap` and the `hold` because of the visual overload. More detail about the family scores in both states and for both *static 1st* and *dynamic 1st* orders is available in the Annex B Figure 15.

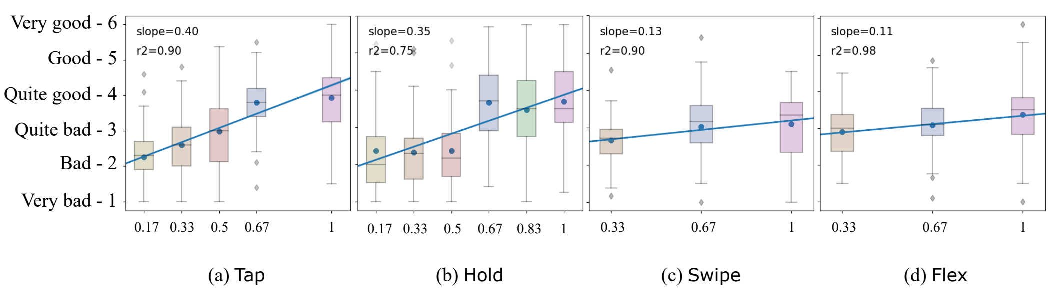


Fig. 8. Distribution of average scores obtained by the representations according to their explicitness.

4.5.3 Quantity of information represented. Based on Figure 8, **families with high explicitness tend to be more appreciated.** This tendency seems to be sharper for `taps` and `holds` than for `swipes` and `flexes` as represented by the slopes of the linear regressions. Nevertheless, we observed one exception, the $_{MatchingShapes}$ family, which is well liked for both `tap` (Mdn = 4.00, CI = [3.50, 4.00], M = 3.61, SD = 1.33, N = 45) and `hold` (Mdn = 3.50, CI = [3.00, 4.00], M = 3.63, SD = 1.32, N = 45), despite its relatively low explicitness (0.33).

Finally, both `taps` ($_t$▼($_i$●);$_t$▲) and `holds` ($_t$▼($_i$●);$_t$■) have a post-touch movement, respectively a release (`tap`) and a pause (`hold`). Families that represent this post-touch movement with separate visual cues performed slightly better (Mdn = 3.50, CI = [3.50, 3.50], M = 3.43, SD = 0.83, N = 45) for the `tap` microgesture than those using superimposed visual cues (Mdn = 2.50, CI = [2.50, 3.00], M = 2.93, SD = 0.72, N = 45).

5 EXPERIMENT 2: REPRESENTING MICROGESTURES ON A HEAD-MOUNTED DISPLAY

The previous study evaluated the perceived efficiencies to graphically represent microgestures on a computer screen with our 21 families, and provided insights on which would work best. Microgestures are frequently deemed promising for Augmented Reality [22, 51, 68, 73], an interactive context relying on head-mounted displays with characteristics that may influence the perceived efficiency of microgestures. Such displays offer interesting opportunities such as graphically representing the microgesture directly on the user's hand. Therefore, in order to refine the results of our first study, we conducted a qualitative controlled laboratory study using an Hololens 2 Augmented Reality headset.

Running an experiment in AR without risking to cause discomfort to participants requires to keep its duration relatively short, typically around 30 minutes [36, 42]. Therefore, we decided to reduce the number of families to test. We chose a subset of 4 families that had both the best scores and distinct designs among the 8 bests families of the online experiment : *Chronophotography*, *ArrowAndBall*, *Chevrons*, *MatchingShapes*. Inspired by LightGuide [58] and Physio@Home [61] we chose to use an in-context approach and mapped our visual cues directly onto the participant's hand as visible on Figure 9

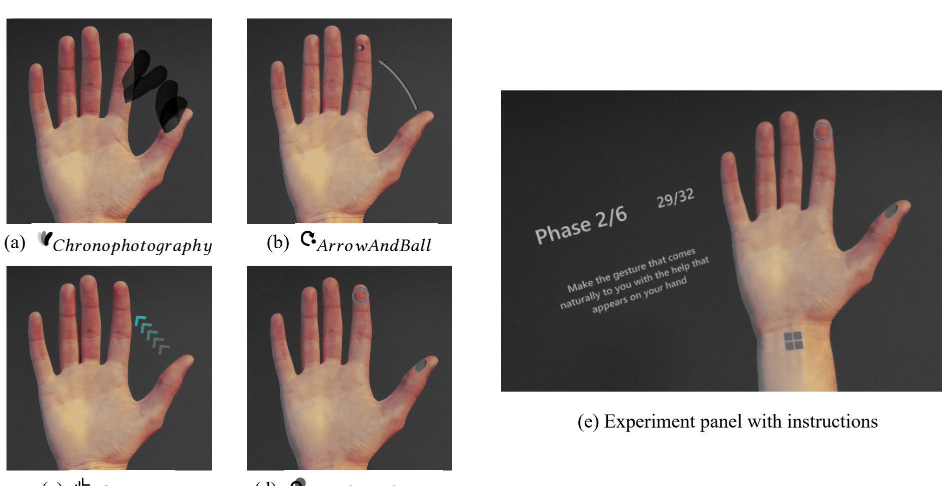
In addition and as opposed to the previous study, we also verify that our representations are correctly interpreted by the participants before asking them to evaluate them. We deliberately chose not to associate a task with each microgesture. The addition of tasks could distort the results since participants would then evaluate both the microgesture representation and the associated task. Therefore, associating a microgesture with a task is outside the scope of this study. We thus allowed participants to focus solely on microgesture representations.

5.1 Participants

We recruited 16 (8 males, 8 females) unpaid volunteers whom age ranged from 15 to 49 years old (M=26, SD=8). We recruited them in our local university. We wanted to gather feedback from participants who could compare representations of the online experience with their equivalents used in this headset experience. As a result, half of them had participated in the online experiment which took place 3 months prior.

5.2 Apparatus

We used a Microsoft Hololens 2 Augmented Reality headset with a wireless connection to the experimenter's computer running the experiment code.



(a) *Chronophotography* (b) *ArrowAndBall* (c) *Chevrons* (d) *MatchingShapes* (e) Experiment panel with instructions

Fig. 9. Tap representations of the 4 chosen families and experiment panel as viewed by a participant wearing the Hololens. The virtual fingers in (a) appear black due to the Hololens capture.

Participants were seated in front of a grey background and were provided with a keyboard and an arm support. Light exposure was controlled so that the experimental conditions were similar for all participants. Participants had a virtual panel displayed on the left side of their right hand, which they could make visible by opening their hand with the palm in front of them. Figure 9e shows the panel indicating the current phase of the experiment along with the corresponding instructions.

5.3 Procedure

This experiment followed a 6-steps procedure:

(1) *Introduction phase:* Participants were introduced with the goals of the study without detailing the microgestures we were interested in. They then answered a short demographic survey. They were also asked if they had participated in the online experiment.
(2) *A priori phase:* Participants were sequentially presented with all microgestures representations in a random order, overlaid on their right hand, and had to execute the microgesture each representation inspired them[5] (without any prior knowledge on the available gestures).
(3) *Scoring phase:* We then introduced the 4 microgestures (`tap`, `swipe`, `flex` and `hold`) and asked participants to evaluate all the microgesture representations on a 6-point Likert scale following the same protocol as the online experiment described previously (i.e. ordered by states and microgestures). Due to the AR headset context, we had to present the representations one by one before letting the participants rate them. The families order was randomized.
(4) *A posteriori phase:* Participants went through the *a priori* phase again, with the same representation order, but this time they had a prior knowledge of the representations and the microgestures they were supposed to represent.
(5) *Subjective phase:* Participants were asked if they had any preference between *static* and *dynamic* representations and if they identified any family or visual cues that they preferred.
(6) *Final discussion phase:* Participants could browse and review the representations as they wanted. This optional final discussion phase allowed them to explore the representations and comment or explain their previous answers.

In total, the experiment lasted between 20 and 45 minutes per participant, variability coming mostly from difference in terms of length of the final discussion phases.

5.4 Results analysis

In this section, we use the same 95% CI method as presented in the previous analysis section.

Because we recruited participants who also participated in the online experiment, we first investigated whether this had an impact on our results. We reviewed all the CIs issued from the combinations of our factors, i.e. the static or dynamic state, the microgesture to recognize and the family, to test if our data reflected a meaningful difference between those who participated in the online experiment and those who did not. For all phases, the scores obtained did not reflect such a difference. For the sake of clarity and conciseness, the plots to validate this statement are provided in the Annex D. Consequently, in the following we consider participants altogether regardless of their participation to the online experiment.

5.4.1 Families preference (Phase 3). The order of presentation of the states (*static 1st* or *dynamic 1st*) did not have an impact in this study. Nevertheless, the state of the representations (*static* or *dynamic*) had an importance in the phase 3 as depicted by Figure 10: overall, we observe a

[5]Since the selected microgestures were clearly distinct and our setup required the presence of an experimenter, we chose not to use a recognizing system that could have caused errors, but to rely on the experimenter's observation.

medium effect size in favor of *dynamic* representations (d>0.5). This effect is mainly due to the $_{Chronophotography}$ family whose *static* representations are disliked by 5 out of 16 participants. Two of them used the word "awful" to describe the $_{Chp}$ family. 3 others had mixed feelings about its *static* representations. However, *dynamic* representations of the $_{Chp}$ family obtains higher results than the *static* ones with a very large effect size (d>1.5). Both $_{ArrowAndBall}$ and $_{Chronophotography}$ families seems to be more suited to *dynamically* represent microgestures than $_{MatchingShapes}$ and $_{Chevrons}$ families (d>0.5).

5.4.2 Microgesture recognition performances (Phases 2 and 4). In this section, we report the average recognition performance across participants. We therefore only report mean values.

In Figure 11, we distinguish the *a priori* (plain) and *a posteriori* (hatched) phases.

In the *a priori* phase, `taps` were recognized 80.47% of the time (M = 80.47%, SD = 24.99, CI = ± 12.25, N = 16) between 11 and 14 times out of 16 participants depending on the families. The $_{Chronophotography}$ and $_{Chevrons}$ *dynamic* representations were slightly better recognized (M = 87.5%, SD = 33.07, CI = ± 16.2, N = 16 for both) than their *static* representations counterparts (resp. M = 68.75%, SD = 46.35, CI = ± 22.71, N = 16 and M = 81.25%, SD = 39.03, CI = ± 19.12, N = 16). The `hold` is however almost never recognized in the *a priori* phase (general M = 12.5%, SD = 18.22, CI = ± 8.93, N = 16, static only M = 12.5%, SD = 17.68, CI = ± 8.66, N = 16, dynamic only M = 12.5%, SD = 21.65, CI = ± 10.61, N = 16). 11 to 14 out of 16 participants executed a `tap` instead. Interestingly, *static* `swipe` representations are *a priori* mistaken with `flexes` 46.88% of the time (M = 46.88%, SD = 36.31, CI = ± 17.79, N = 16). Indeed, depending on the family, 5 to 7 participants recognized *static* `swipe` representations as flexions of the proximal interphalangeal[6] joint of the finger. We observe the same behavior for the *dynamic* state (M = 40.62%, SD = 31.72, CI = ± 15.54, N = 16) except for the $_{Chp}$ family whose *dynamic* representation is correctly recognized 81.25% of the time (M = 81.25%, SD = 39.03, CI = ± 19.12, N = 16). Finally, `flexes` were correctly *a priori* recognized by 91.41% of the time (M = 91.41%, SD = 16.37, CI = ± 8.02, N = 16). The few participants that mixed `swipes` with `flexes` purposefully performed flexions of the metacarpophalangeal joint for `flexes` representations.

In the *a posteriori* phase, thus after the *scoring* phase and therefore having acquired knowledge on the microgestures our representations were associated to, *dynamic* `taps` were correctly recognized 89.06% of the time (M = 89.06%, SD = 26.47, CI = ± 12.97, N = 16). `Holds` were better recognized in this phase, while still being far from perfect (general M = 71.09%, SD = 18.07, CI = ± 8.85, N = 16,

[6] See *DIP, PIP and MCP joints of the hand* [Wikipedia]

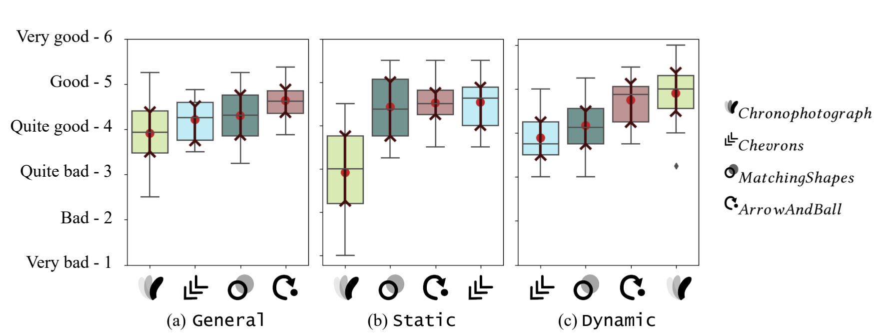


Fig. 10. Reported quality scores of the representations with their respective quartiles, mean (red dot) and their 95% bootstrapped CI for the median per family.

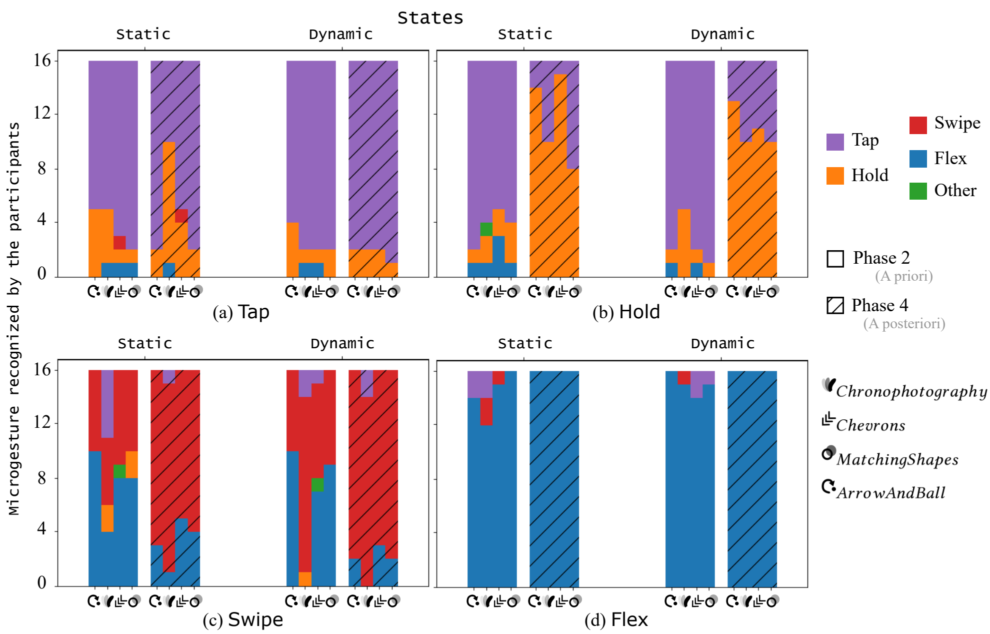


Fig. 11. Clustered stacked bars of recognized microgestures arranged by family, state and phase. The more a bar is monochrome, the more a representation has been recognized in the same way by all the participants.

static only M = 73.44%, SD = 20.67, CI = ± 10.13, N = 16, dynamic only M = 68.75%, SD = 29.67, CI = ± 14.69, N = 16). $_{Chronophotography}$ and $_{MatchingShapes}$ were the most difficult families to recognize (resp. M = 62.5%, SD = 33.07, CI = ± 16.2, N = 16 and M = 56.25%, SD = 34.8, CI = ± 17.05, N = 16). For the $_{Chp}$ family, participants did not notice the color difference between `tap` and `hold` representations. `Swipe` representations were *a posteriori* better understood (from M = 44.56%, SD = 33.94, CI = ± 16.63, N = 16 to M = 82.03%, SD = 25.76, CI = ± 12.62, N = 16) but 1 to 3 participants continued to confuse them with `flexes`. We also observed that `flexes` are perfectly recognized *a posteriori*.

5.4.3 Participants subjective feedback. The exchanges with participants during the final discussion phase revealed interesting extreme positioning regarding their preference of representation. This is especially salient with the $_{Chronophotography}$ family. Beyond their preference, 8 out of 16 participants explained that the $_{Chp}$ family, based on a virtual finger, is easier to understand with *dynamic* visual cues. However, 4 out of 16 participants stated that the $_{Chp}$ family implies a heavy cognitive load. They felt "overload[ed] like crazy" or "[could not] see anything", regardless of whether the representation was *static* or *dynamic*.

The $_{MatchingShapes}$ family is based on the metaphor of filling a hole with a disk, as can be observed in some early games [72] as well as in the μGlyph notation [10, 11]. 7 out of 16 participants reported to have understood this metaphor. The other participants stated that they preferred an arrow-based representation over the $_{MaS}$ representation. In addition, the $_{MaS}$ family implemented the metaphor of a stop road sign, i.e. a hexagon to indicate to stop, for the `hold`, i.e. touch and stop. However, none of the participants understood it, which emphasizes the difficulty to rely on metaphors. Indeed, **a metaphor can be missed without all its visual associations**, i.e octogon shape plus red sign plus "Stop" text for the stop road sign. 7 out of 16 participants indicated that they better understood the distinction between `hold` and `tap` using different sizes: `holds` being conveyed by a large sphere for $_{ArrowAndBall}$ or a large ending arrow for $_{Chevrons}$ in opposition with `taps` conveyed by a small sphere or arrow.

Although we did not observe a strong impact of dynamicity on the representations recognition, 9 participants out of 16 indicated that they had a clear **preference for dynamic representations**. The other were more mitigated and their preferences varied depending on the microgesture and the family. Moreover 13 participants felt that dynamic representations are essential to understand the distinction between `hold` and `tap`.

Finally, our experiment used visual cues directly positioned in relation to the participants' hands. However, very few participants tried on their own to rotate their hand to get a better understanding of some `flexes` representations during the *a priori* and *a posteriori* phases One of our participants explained his preference for the $_{A\&B}$ family because they could better understand it without having to move the wrist. As they suggested, we should thus conceive `flex` representations directly understandable with a frontal point of view.

6 DISCUSSION

The following guidelines have been first issued from the quantitative data of the online experiment and have been refined and formulated based on the qualitative analysis of the headset experiment. Since we purposefully designed our experimental protocols to represent one microgesture at a time, these guidelines only apply when separately representing microgestures. If one wants to show multiple microgestures simultaneously, new studies should first be carried out, as other factors such as visual clutter could come into play. We observe for the representation of a single microgesture that the visual load varies from one family to another. It is therefore mandatory to consider what is important to represent first.

6.1 *What* to represent of a microgesture?

DG1 - Implicitly or explicitly, show the actuator - Despite the fact that we indicated to the participants that we are studying thumb-to-finger microgestures, we observe that every representation that do not show the actuator, e.g. $_{Ring}$, $_{EmphasisLines}$ or $_{HighlightedZone}$, obtain strictly negative results whereas representations showing it explicitly, e.g. $_{MatchingShapes}$ and $_{ElectricCharges}$, or implicitly, e.g. $_{ArrowAndBall}$ and $_{Chevrons}$, obtain better results. Consequently, results suggest that only showing the receiver is not sufficient to depict a movement even if the related actuator has been declared beforehand.

DG2 - Prefer to maximize the given information - Beyond the actuator, our results show that the more information is given to represent a microgesture the better. For instance, the $_{DoublePathArrow}$ family shows every piece of information and is preferred to the $_{MovementStroke}$ which omits the movement direction.

DG3 - Show the trajectory and/or the direction if possible - For cases where displayed information has to be limited, it is preferable to clearly show the trajectory of a microgesture. Indeed, we observe that both the actuator and the receiver can be shown implicitly through both ends of a visual cue representing the trajectory. Similarly, explicitly representing the direction clears out which finger is the actuator and which is the receiver. This is shown by the $_{ArrowAndBall}$ family whose arrow and line describe a movement from one finger onto the other. Looking at the results of the $_{ArrowTip}$ family that are inconsistent among microgestures, it seems that showing the microgesture direction alone should be avoided.

DG4 - Describe each step with separate cues - According to the results of the online experiment, families that show the trajectory tend to be preferred when using separate visual cues to represent each step, e.g. $_{BrokenArrow}$ or $_{DoublePathArrow}$, compared to those using superimposed visual cues, e.g. $_{MovementStroke}$ or $_{DashedBalls}$.

6.2 *How* to represent our chosen *What*?

DG5 - Prefer arrows or partially transparent fingers - Our results suggest that using arrows with a line is a good practice which allows great flexibility in design. Indeed, all 4 families using arrows with a line are among our top 8. Using stroboscopic finger shapes, i.e. $_{Chronophotography}$ family, also seems to be an efficient method but for *dynamic* representations only. However, based on certain participants' comments ("I deeply hate the [$_{Chronophotography}$ family]", "it overloads like crazy", "I cannot see anything"), it already seems that the $_{Chp}$ family might not be suited to represent multiple microgestures at once due to its high visual load.

DG6 - Be careful with visual associations - For context-switching microgestures, representing the actuator and the receiver with a pair of visual cues to suggest the microgesture without explicitly showing its trajectory can yield opposite results. For instance, the $_{ElectricCharges}$ family uses paired color and text cues, i.e 'minus' blue and 'plus' red symbols, and obtained an average results in the online experiment, while the $_{MatchingShapes}$ family uses empty/filled matching shapes and is one of the highest rated families. However, while still scoring positively in the headset experiment, the $_{MaS}$ family has been judged less effective for conveying the distinction between the `tap` *release*, using round shapes, and the `hold` *pause*, using octagon shapes. Our participants better recalled this distinction using different sizes, as done by $_{ArrowAndBall}$ spheres or $_{Chevrons}$ arrows. Thus, the use of visual associations appears to be a promising technique but it requires preliminary testing/studies.

DG7 - Use animations to increase satisfaction, but it is not needed to increase recognition - Even though our participants expressed their preference for *dynamic* representations, their scores did not reflect it. Furthermore, Figure 11 of the headset experiment indicates that the animation did not have a significant impact on recognition except for the $_{Chronophotography}$ family. As in **DG5**, participants explained this difference by the visual overload of the $_{Chp}$ *static* design.

6.3 3D-specific guideline

DG8 - Prefer to use representations understandable with any hand posture - In case a user have to maintain their hand or a virtual one in a specific position/orientation, the microgesture representations should remain understandable. This has been highlighted by the participants in the headset experiment ("[$_{ArrowAndBall}$] is better because visually it is wider in front of you", "Otherwise, you are forced to turn your hand") for the `flex` microgesture for which the $_{ArrowAndBall}$ family has been considered as the only one understandable with any hand posture.

6.4 Comparative discussion to existing approaches and representations

The taxonomy established by Antoine *et al.* [2] has been used as a basis for the creation of the 21 families studied (see Section 3.1). We enrich this taxonomic work, which is only descriptive and not normative, with guidelines. Our guidelines are consistent with McAweeney et al.'s [48] observations on how to represent the *motion* and *touch* aspects of a gesture. Indeed, McAweeney et al. observed that their 30 end-user participants spontaneously drew arrows and contact points to represent the movement of the gesture ("Movement was represented [...] by dots or arrows on the joints") and touch ("Touch was [...] communicated by contact points"). These observations are consistent with guidelines DG1, DG3 and DG5. More than concrete representations proposed by end-users, we provide guidelines that can help to create or improve representations of microgestures.

We detail two examples of how our guidelines along with the 21 created families and the explicitness metric can be applied to improve existing microgesture representations. Figure 12 presents two microgesture representations that are reworked versions of existing representations obtained by applying our guidelines. The initial representations from the literature contained missing or redundant information.

The representation of Figure 12a lacks information producing an ambiguous representation that could be mistaken for a `tap`, a `hold` or even a `press`. In the original paper, the author chose to use a textual description to avoid this ambiguity. There is a consensus word to describe the `tap` microgesture, but Chaffangeon Caillet *et al.* explained that for other microgestures "the names [do] not give a clear idea of the corresponding microgestures" [11]. They also explained that readers tended to rely "solely on the associated images," which highlights the importance of our work on the visual representation of microgestures. Figure 12a* maximizes the given information (DG2) to disambiguate the existing representation and uses both arrows (DG3, DG5) and colored fingers to highlight the receiver and the actuator (DG1).

Figure 12b shows redundant information about a microgesture: indeed, both the colored fingers and the chronophotograph emphasize the actuator and the receiver. This representation follows one of the observations given by McAweeney *et al.*: the chronophotography should be used as it "shows a clear beginning and end position". However, our studies suggested that the superimposition of multiple transparent fingers produced more confusion and was not better than using arrows for *static* representations (DG5). We thus reused the $\curvearrowleft_{BA}$ family in our proposed redesign in Figure 12b*.

In general, we can summarize the previous examples with the following method:

(1) **List the explicitness features**: list all the explicitness features of the considered microgesture as introduced by Figure 6 (trajectory, direction, actuator, receiver, release, separate for `taps`; trajectory, direction, actuator, receiver, pause, separate for `holds`; trajectory, direction, explicit touch point for `swipes`; trajectory, direction, explicit moving part for `flexes`).
(2) **Assess each feature**: for each feature assess whether the representation is correct, incomplete, ambiguous or redundant.

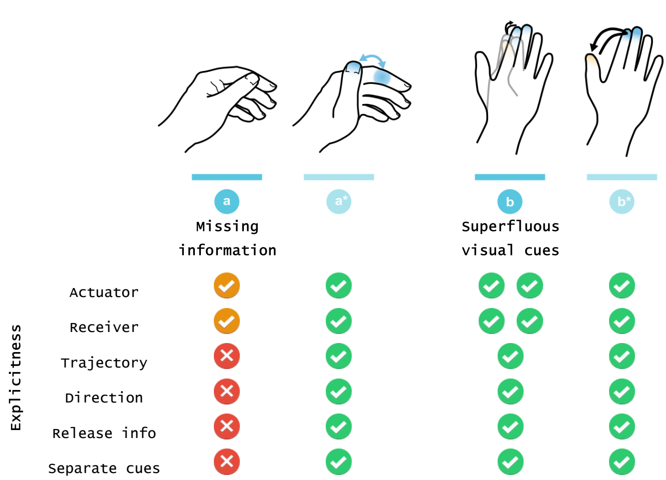


Fig. 12. Tap illustrations inspired from papers and their reworked version (marked by a *). Each illustration is associated to its explicitness decomposition with green ticks for explicit information, orange ticks for implicit information and red crosses for missing information. (a) inspired from DigiTouch [68], (b) inspired from µGlyph [11].

(3) **Correct each feature by applying the guidelines**: for each feature change the representation accordingly (e.g. add information to an incomplete representation, make information explicit in an ambiguous representation, remove pieces of information if they are redundant).

To our knowledge, this method is the first providing a systematic approach to enhance microgesture representations.

6.5 Families to consider for practitioners

Practitioners may want to use only certain microgestures for a given system, e.g. only taps and swipes. To this end, we summarize in Tables 3 and 4 the top-3 best scoring families from the online experiment for each possible subset of microgestures and each states. In addition, we did not observe any influence of the state order, i.e. static 1st or dynamic 1st, nor of the order of appearance of the microgestures on the results of our study. Thus, it seems that `tap`, `hold`, `flex` and `swipe` microgestures can be shown in any sequential order when introducing the functionalities of a system, which remains true regardless of the dynamism of the representations.

State	`tap`	`swipe`	`flex`	`hold`	Top 1	Top 2	Top 3
static	x	x	x	x	*A&B* (4.02)	*Chv* (3.92)	*DpA* (3.88)
	x	x	x		*DpA* (3.94)	*A&B* (3.9)	*Chv* (3.89)
	x	x		x	*A&B* (4.03)	*DpA* (3.82)	*Chv* (3.79)
	x	x			*DpA* (3.88)	*A&B* (3.86)	*Chv* (3.68)
	x		x	x	*A&B* (4.21)	*Chv* (4.02)	*DpA* (4.01)
	x		x		*DpA* (4.17)	*A&B* (4.13)	*Chv* (4.03)
	x			x	*A&B* (4.33)	*DpA* (3.99)	*MaS* (3.93)
	x				*A&B* (4.29)	*DpA* (4.27)	*MaS* (3.96)
		x	x	x	*Chv* (3.97)	*A&B* (3.93)	*DpA* (3.76)
		x	x		*Chp* (4.01)	*Chv* (3.96)	*DpA* (3.78)
		x		x	*A&B* (3.9)	*Chv* (3.8)	*DpA* (3.6)
		x			*CwS* (3.64)	*Chp* (3.62)	*Chv* (3.6)
			x	x	*A&B* (4.18)	*Chv* (4.16)	*DpA* (3.89)
			x		*Chp* (4.4)	*Chv* (4.31)	*DpA* (4.07)
				x	*A&B* (4.38)	*Chv* (4.0)	*MaS* (3.91)
dynamic	x	x	x	x	*Chp* (4.9)	*Chv* (3.76)	*MSt* (3.74)
	x	x	x		*Chp* (4.98)	*MaS* (3.71)	*Chv* (3.64)
	x	x		x	*Chp* (4.87)	*MSt* (3.73)	*Chv* (3.65)
	x	x			*Chp* (4.97)	*MaS* (3.46)	*DpA* (3.43)
	x		x	x	*Chp* (4.93)	*Chv* (3.99)	*MSt* (3.95)
	x		x		*Chp* (5.07)	*BA* (3.92)	*Chv* (3.92)
	x			x	*Chp* (4.9)	*MSt* (4.03)	*BA* (3.98)
	x				*Chp* (5.13)	*BA* (4.02)	*DpA* (3.91)
		x	x	x	*Chp* (4.82)	*MSt* (3.81)	*Chv* (3.75)
		x	x		*Chp* (4.9)	*MaS* (3.93)	*Chv* (3.57)
		x		x	*Chp* (4.73)	*MSt* (3.82)	*Chv* (3.59)
		x			*Chp* (4.8)	*MaS* (3.64)	*CwS* (3.47)
			x	x	*Chp* (4.83)	*MSt* (4.16)	*Chv* (4.09)
			x		*Chp* (5.0)	*MaS* (4.22)	*CnS* (4.16)
				x	*Chp* (4.67)	*MSt* (4.53)	*Chv* (4.11)

Table 3. The 3 best representations with their mean score for each subset of the `tap`, `hold`, `swipe` and `flex` microgestures meant to be used in their *static* **or** *dynamic* state (based on the results of our online experiment).

State	tap	swipe	flex	hold	Top 1	Top 2	Top 3
static & dynamic	x	x	x	x	*Chp* (4.28)	*Chv* (3.84)	*A&B* (3.74)
	x	x	x		*Chp* (4.4)	*Chv* (3.76)	*DpA* (3.68)
	x	x		x	*Chp* (4.14)	*A&B* (3.76)	*Chv* (3.72)
	x	x			*Chp* (4.26)	*DpA* (3.66)	*A&B* (3.64)
	x		x	x	*Chp* (4.3)	*Chv* (4.0)	*A&B* (3.88)
	x		x		*Chp* (4.5)	*Chv* (3.98)	*DpA* (3.91)
	x			x	*Chp* (4.1)	*A&B* (3.98)	*Chv* (3.91)
	x				*Chp* (4.3)	*DpA* (4.09)	*A&B* (3.96)
		x	x	x	*Chp* (4.27)	*Chv* (3.86)	*A&B* (3.67)
		x	x		*Chp* (4.46)	*Chv* (3.76)	*MaS* (3.6)
		x		x	*Chp* (4.06)	*Chv* (3.69)	*A&B* (3.66)
		x			*Chp* (4.21)	*CwS* (3.56)	*Chv* (3.33)
			x	x	*Chp* (4.3)	*Chv* (4.12)	*A&B* (3.84)
			x		*Chp* (4.7)	*Chv* (4.19)	*MaS* (3.91)
				x	*Chv* (4.06)	*MSt* (4.01)	*A&B* (4.0)

Table 4. The 3 best representations with their mean score for each subset of the tap, hold, swipe and flex microgestures meant to be used in their *static* **and** *dynamic* state (based on the results of our online experiment).

7 FUTURE WORK

In this work, we examined an informed selection of visual representations of microgestures which are part of a large design space. In order to lay the ground for best representing microgestures, we aimed at identifying first insights on what elements of a microgesture should be represented and how to represent them. We conducted two empirical studies allowing us to observe and report design guidelines, yet more comprehensive and controlled experiments are clearly needed. In particular and based on our results, one logical next step is to conduct controlled experiment to further investigate visual representations of composed microgestures, i.e. microgestures composed of multiple elementary event as defined by Chaffangeon Caillet *et al.* [10, 11], which can share movement characteristics and therefore lead to ambiguities. Another logical step, is to study the representation of multiple microgestures presented at once, which will arguably be a widespread use case, e.g. systems offering multiple possible commands at any given times. We designed our representations through a single point of view, however studying the impact of the angle at which a representation is seen should be investigated.

Composed microgestures - Our work already studied an ambiguous case with tap $_{t}\overset{\circ\bullet}{\blacktriangledown}(_{i}\bullet);_{t}\overset{\bullet\circ}{\blacktriangle}$ and hold $_{t}\overset{\circ\bullet}{\blacktriangledown}(_{i}\bullet);_{t}\overset{\bullet}{\blacksquare}$, both sharing the same initial *touch* event $_{t}\overset{\circ\bullet}{\blacktriangledown}(_{i}\bullet)$ while differing in their respective final event (tap: *release* $_{t}\overset{\bullet\circ}{\blacktriangle}$; hold: *pause* $_{t}\overset{\bullet}{\blacksquare}$). It seems that the strategy used to represent several events of a microgesture, i.e. superimposing the representation of its successive events or representing each events separately, has a significant impact on the understandability of a representation. Further controlled studies should focus on identifying the best strategies.

Multiple microgestures - Our work only focuses on representing one microgesture on the hand at a time to avoid potential interaction effect between representations. However, multiple microgestures are sometimes represented in the same picture [59]. At the very least, two cases exist: representing the same microgesture but performed with different actuators and/or receivers, e.g. a tap of the index or middle finger on the thumb or the palm; and representing different microgesture types involving the same actuator and/or receiver, e.g. tap or swipe of the thumb on the index. Considering several different microgestures at once on the hand could therefore have an impact on how to represent them most effectively.

Viewpoint - Our work on microgesture representations uses an opened hand as initial posture. This choice was based on existing systems, such as the Hololens 2 which triggers the apparition of a menu when the hand is raised into the user's field of view. Our rational was therefore to use a similar trigger for a microgesture help system. However, the intelligibility of the representations may vary when the viewpoint can change dynamically or, in static cases, when they are simply presented with a different hand posture. Consequently, further studies should focus on the impact of viewpoint and hand posture on microgesture representations.

Looking at the big picture, the representation of **composed** microgestures and **multiple** microgestures can be studied together. For instance, `tap` and `hold` microgestures which share their initial movement, could lead to merged representations showing both possibilities at once. It could potentially avoid visual clutter, as done in previous work such as Arpege [28] or Octopocus [7]. To remain consistent with the literature, we also chose to represent the `swipe` only with its middle `drag` $_t\ddot{\blacktriangledown}(_i\bullet)\,|\,_t\ddot{\blacktriangle}(_i\bullet)$ event, which show two possible directions, i.e. **multiple** microgestures, but not several μGlyph events, i.e. **composed** microgestures. Consequently, we have only scratched the surface of this design approach as the `swipe` motion implicitly starts with a *touch* event: $_t\overset{\circ\bullet}{\blacktriangledown}(_i\bullet);(_t\ddot{\blacktriangledown}(_i\bullet)\,|\,_t\ddot{\blacktriangle}(_i\bullet));_t\overset{\bullet\circ}{\blacktriangle}$.

Finally, designers commonly use videos that include hand animation, e.g. Kinect[7], MagicalHands [4], and StrikeAPose [65]. These video representations can also use *static* or *dynamic* visual cues along with a moving hand. Inspired by these examples, we initially thought about integrating the movement of the hand itself as another variable of our experiment. However, doing so could have brought confusion between what is due to the animation of the visual cue and what is due to the animation of the hand. Consequently, for our *dynamic* representations, we only animated the visual cues but not the hand itself. Nevertheless, we believe that most of the families can easily be adapted to video formats. For instance, $_{ArrowAndBall}$, $_{BrokenArrow}$ and $_{DoublePathArrow}$ families only need to adopt a responsive behavior with respect to the hand posture. In future studies, we would like to use the concept of families with a moving hand, both in a video format and in Augmented Reality, with visual cues that would adopt a responsive behavior with respect to the user's hand.

8 CONCLUSION

This paper focuses on single-image representations of commonly used microgestures. As a first attempt to explore the large design space, we conducted an online experiment and a controlled AR experiment to compare families of microgesture representations. We aimed at understanding *what* should be represented of a microgesture and *how* to depict this information. The 21 families created shaped our contribution and our experimental results led us to establish design guidelines and therefore first insights on best representation practices. In particular, we recommend to explicitly or implicitly represent actuators, movement trajectory and directions of a microgesture, and to use dynamic representations for user engagement. These findings can be used by researchers and practitioners to represent microgestures. The guidelines, along with the discussion of next steps that the microgesture community should focus on to expand them, lay the foundation for future work on microgesture representation and guidance, a critical aspect for the adoption of microgesture interaction.

ACKNOWLEDGMENTS

This work was partly supported by the French National Research Agency (ANR) project MIC (ANR-22-CE33-0017) and project Discovery (ANR-19-CE33-0006).

[7]See this tutorial with *static* and *dynamic* body gesture annotations [Kinect tutorial]

REFERENCES

[1] Yoshizawa Akira. *Beautiful Origami / Utsukushi origami*. Kamakura Shobo, January 1974.

[2] 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*, May 2021.

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

[4] Rahul Arora, Rubaiat Habib Kazi, Danny M. Kaufman, Wilmot Li, and Karan Singh. MagicalHands: Mid-Air Hand Gestures for Animating in VR. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 463–477, New Orleans LA USA, October 2019. ACM.

[5] Christopher Austin, Barrett Ens, Kadek Satriadi, and Bernhard Jenny. Elicitation study investigating hand and foot gesture interaction for immersive maps in augmented reality. *Cartography and Geographic Information Science*, 47:1–15, January 2020.

[6] Ronald Baecker, Ian Small, and Richard Mander. Bringing icons to life. In *Proceedings of the SIGCHI conference on Human factors in computing systems Reaching through technology - CHI '91*, pages 1–6, New Orleans, Louisiana, United States, 1991. ACM Press.

[7] Olivier Bau and Wendy E. Mackay. OctoPocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology - UIST '08*, page 37, Monterey, CA, USA, 2008. ACM Press.

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

[9] 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*, pages 150–157. Association for Computing Machinery, October 2018.

[10] 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)*, April 2022.

[11] Adrien Chaffangeon Caillet, Alix Goguey, and Laurence Nigay. μGlyph: a Microgesture Notation. 2023.

[12] Amira Chalbi, Jacob Ritchie, Deokgun Park, Jungu Choi, Nicolas Roussel, Niklas Elmqvist, and Fanny Chevalier. Common Fate for Animated Transitions in Visualization. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2019.

[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, Yi-Ling Chen, Chi-Hao Hsieh, Rong-Hao Liang, and Bing-Yu Chen. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 549–556, Charlotte NC USA, November 2015. ACM.

[15] 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.

[16] Ulysse Cote Allard, Francois Nougarou, Cheikh Latyr Fall, Philippe Giguere, Clement Gosselin, Francois Laviolette, and Benoit Gosselin. A convolutional neural network for robotic arm guidance using sEMG based frequency-features. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2464–2470, Daejeon, South Korea, October 2016. IEEE.

[17] James E Cutting. Representing Motion in a Static Image: Constraints and Parallels in Art, Science, and Popular Culture. *Perception*, 31(10):1165–1193, October 2002. Number: 10.

[18] Thomas De Fanti and Daniel Sandin. Sayre Glove. Technical report, University of Illinois, Chicago Circle, Box 4348 Chicago Il 60680, 1977.

[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] Christoph Endres, Tim Schwartz, and Christian A. Müller. Geremin": 2D microgestures for drivers based on electric field sensing. In *Proceedings of the 16th international conference on Intelligent user interfaces*, IUI '11, pages 327–330, New York, NY, USA, 2011. Association for Computing Machinery.

[22] Barrett Ens, Aaron Quigley, Hui-Shyong Yeo, Pourang Irani, Thammathip Piumsomboon, and Mark Billinghurst. Counterpoint: Exploring Mixed-Scale Gesture Interaction for AR Applications. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6, Montreal QC Canada, April 2018. ACM.

[23] Jacqui Fashimpaur, Kenrick Kin, and Matt Longest. PinchType: Text Entry for Virtual and Augmented Reality Using Comfortable Thumb to Fingertip Pinches. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–7, Honolulu HI USA, April 2020. ACM.
[24] Euan Freeman, Gareth Griffiths, and Stephen A. Brewster. Rhythmic micro-gestures: discreet interaction on-the-go. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 115–119, Glasgow UK, November 2017. ACM.
[25] Asaf Friedman. The Role of Visual Design in Game Design. *Games and Culture*, 10(3):291–305, May 2015. Number: 3.
[26] Roman Ganhör and Wolfgang Spreicer. Monox: extensible gesture notation for mobile devices. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, pages 203–212, Toronto, ON, Canada, 2014. ACM Press.
[27] Huaiying Gao. *The Effects of Still Images and Animated Images on.* PhD thesis, Virginia Polytechnic Institute and State University, 2005.
[28] Emilien Ghomi, Stéphane Huot, Olivier Bau, Michel Beaudouin-Lafon, and Wendy E. Mackay. Arpège: learning multitouch chord gestures vocabularies. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, pages 209–218, St. Andrews Scotland, United Kingdom, October 2013. ACM.
[29] Emmanouil Giannisakis, Gilles Bailly, Sylvain Malacria, and Fanny Chevalier. IconHK: Using Toolbar button Icons to Communicate Keyboard Shortcuts. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 4715–4726, New York, NY, USA, 2017. Association for Computing Machinery.
[30] Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pages 553–563, New York, NY, USA, October 2017. Association for Computing Machinery.
[31] Jeffrey T. Hansberger, Chao Peng, Shannon L. Mathis, Vaidyanath Areyur Shanthakumar, Sarah C. Meacham, Lizhou Cao, and Victoria R. Blakely. Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions. In *Virtual, Augmented and Mixed Reality*, volume 10280, pages 505–520. Springer International Publishing, Cham, 2017. Series Title: Lecture Notes in Computer Science.
[32] 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.
[33] Thomas Hull, editor. *Origami^3: Third International Meeting of Origami Science, Math, and Education.* A K Peters, Natick, MA, 2002. Meeting Name: International Meeting of Origami Science, Mathematics, and Education.
[34] Renate Häuslschmid, Benjamin Menrad, and Andreas Butz. Freehand vs. micro gestures in the car: Driving performance and user experience. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 159–160, March 2015.
[35] 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.
[36] Yuan Jing, Mansouri Behzad, Pettey Jeff H, Ahmed Sarah Farukhi, and Khaderi S Khizer. The Visual Effects Associated with Head-Mounted Displays. *International Journal of Ophthalmology and Clinical Research*, 5(2), June 2018. Number: 2.
[37] Kenrick Kin, Björn Hartmann, Tony DeRose, and Maneesh Agrawala. Proton++: a customizable declarative multitouch framework. In *Proceedings of the 25th annual ACM symposium on User interface software and technology - UIST '12*, page 477, Cambridge, Massachusetts, USA, 2012. ACM Press.
[38] Joseph LaViola and Robert Zeleznik. Flex And Pinch: A Case Study Of Whole Hand Input Design For Virtual Environment Interaction. *Brown University Site of the NSF Science and Technology Center for Computer Graphics and Scientific Visualization*, page 5, June 2000.
[39] DoYoung Lee, SooHwan Lee, and Ian Oakley. Nailz: Sensing Hand Input with Touch Sensitive Nails. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA, April 2020. ACM.
[40] 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.
[41] Ziheng Li, Zhenyuan Lei, An Yan, Erin Solovey, and Kaveh Pahlavan. ThuMouse: A Micro-gesture Cursor Input through mmWave Radar-based Interaction. In *2020 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–9, January 2020. ISSN: 2158-4001.
[42] Sungjin Lim, Hosung Jeon, Minwoo Jung, Chulwoong Lee, Woonchan Moon, Kwangsoo Kim, Hwi Kim, and Joonku Hahn. Fatigue-free visual perception of high-density super-multiview augmented reality images. *Scientific Reports*, 12(1):2959, February 2022. Number: 1.
[43] Christian Loclair and Sean Gustafson. PinchWatch: A Wearable Device for One-Handed Microinteractions. *undefined*, 2010.

[44] Eva Mackamul, Géry Casiez, and Sylvain Malacria. Exploring visual signifier characteristics to improve the perception of affordances of in-place touch inputs. *Proc. ACM Hum.-Comput. Interact.*, 7(MHCI), sep 2023.

[45] 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.

[46] Mamaylya. HoloLens 2 gestures (for example, gaze and air tap) for navigating a guide in Dynamics 365 Guides - Dynamics 365 Mixed Reality, November 2022.

[47] Sven Mayer, Lars Lischke, Adrian Lanksweirt, Huy Viet Le, and Niels Henze. How to communicate new input techniques. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, pages 460–472, Oslo Norway, September 2018. ACM.

[48] Erin McAweeney, Haihua Zhang, and Michael Nebeling. User-Driven Design Principles for Gesture Representations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Montreal QC Canada, April 2018. ACM.

[49] Meredith Ringel Morris, Jacob O. Wobbrock, and Andrew D. Wilson. Understanding users' preferences for surface gestures. In *Proceedings of Graphics Interface 2010*, GI '10, pages 261–268, CAN, 2010. Canadian Information Processing Society.

[50] Alex Olwal, Thad Starner, and Gowa Mainini. E-Textile Microinteractions: Augmenting Twist with Flick, Slide and Grasp Gestures for Soft Electronics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA, April 2020. ACM.

[51] 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.

[52] Firas Raheem and Hadeer Raheem. ASL Recognition Quality Analysis Based on Sensory Gloves and MLP Neural Network. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 47:1–20, September 2018.

[53] Jean-Christian Ricard. *Equitation, locomotion et mécanisme des allures au XIX e siècle: De la méthode graphique à la chronophotographie*, volume 41 of *Revue d'histoire des sciences.* Armand Colin, December 1988.

[54] Gustavo Rovelo, Donald Degraen, Davy Vanacken, Kris Luyten, and Karin Coninx. Gestu-Wan - An Intelligible Mid-Air Gesture Guidance System for Walk-up-and-Use Displays. In Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2015*, volume 9297, pages 368–386. Springer International Publishing, Cham, 2015. Series Title: Lecture Notes in Computer Science.

[55] 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.

[56] Adwait Sharma, Joan Sol Roo, and Jürgen Steimle. Grasping Microgestures: Eliciting Single-hand Microgestures for Handheld Objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Glasgow Scotland Uk, May 2019. ACM.

[57] Leonard J. Simms, Kerry Zelazny, Trevor F. Williams, and Lee Bernstein. Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, 31(4):557–566, April 2019.

[58] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 179–188, Austin Texas USA, May 2012. ACM.

[59] 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.

[60] Joachim Stöhr and Hans Christoph Siegmann. *Magnetism: from fundamentals to nanoscale dynamics.* Number 152 in Springer series in solid-state sciences. Springer, Berlin ; New York, 2006. OCLC: ocm72867752.

[61] Richard Tang, Hesam Alizadeh, Anthony Tang, Scott Bateman, and Joaquim A.P. Jorge. Physio@Home: design explorations to support movement guidance. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, pages 1651–1656, Toronto Ontario Canada, April 2014. ACM.

[62] Carl Therrien. *Illusion, idéalisation, gratification : l'immersion dans les univers de fiction à l'ère du jeu vidéo.* PhD thesis, Université du Québec, Montréal, June 2011.

[63] Cyril Thomas, André Didierjean, and Serge Nicolas. Scientific Study of Magic: Binet's Pioneering Approach Based on Observations and Chronophotography. *The American Journal of Psychology*, 129(3):313–326, October 2016. Number: 3.

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

[65] Robert Walter, Gilles Bailly, and Jörg Müller. StrikeAPose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 841–850, Paris France, April 2013. ACM.

[66] 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.

[67] David Way. *EMGRIE: Ergonomic Microgesture Recognition and Interaction Evaluation, A Case Study.* PhD thesis, Massachusetts Institute of Technology, 2014.

[68] 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.

[69] 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.

[70] Zheer Xu, Pui Chung Wong, Jun Gong, Te-Yen Wu, Aditya Shekhar Nittala, Xiaojun Bi, Jürgen Steimle, Hongbo Fu, Kening Zhu, and Xing-Dong Yang. TipText: Eyes-Free Text Entry on a Fingertip Keyboard. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pages 883–899, New York, NY, USA, October 2019. Association for Computing Machinery.

[71] Alexandra Yeung, Siegbert Schmid, Adrian George, and Michael King. Still pictures, animations or interactivity – What is more effective for elearning? *Proceedings of The Australian Conference on Science and Mathematics Education*, 2007.

[72] Jennifer M. Zosh, Brian N. Verdine, Andrew Filipowicz, Roberta Michnick Golinkoff, Kathy Hirsh-Pasek, and Nora S. Newcombe. Talking Shape: Parental Language With Electronic Versus Traditional Shape Sorters: Traditional Toys Promote Parent Spatial Talk. *Mind, Brain, and Education*, 9(3):136–144, September 2015. Number: 3.

[73] Klen Čopič Pucihar, Christian Sandor, Matjaž Kljun, Wolfgang Huerst, Alexander Plopski, Takafumi Taketomi, Hirokazu Kato, and Luis A. Leiva. The Missing Interface: Micro-Gestures on Augmented Objects. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6, Glasgow Scotland Uk, May 2019. ACM.

A COVERING OF THE VISUAL CUES WITH OUR FAMILIES

Each family uses visual cues corresponding to different subsets of the categories introduced by Table 1. Table 5 details the categories corresponding to each microgesture representation of our 21 families.

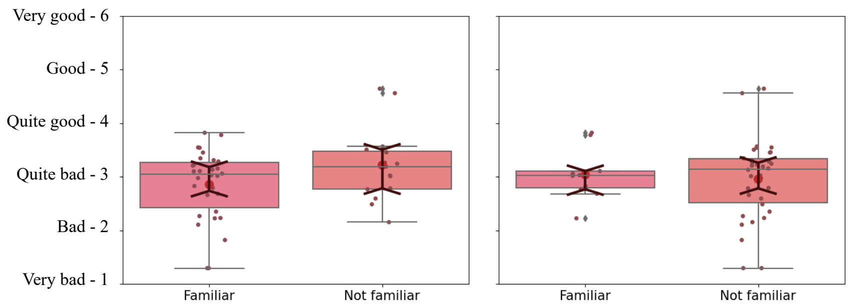
		Trajectory				Direction			Actuator & Receiver			Emphasis tools				
		Lines				Arrows			Shapes			Stroke		Fill		
Family	Microgesture	Simple	Curve	Broken	Stroboscopic effect	Simple-headed	Double-headed	Motion lines	Organic	Geometric	Icon	Dash pattern	Color	Opacity	Color	Text
ArrowAndBall	Tap		x			x				x						
	Swipe	x					x			x						
	Flex		x				x			x						
	Hold		x			x				x						
DoublePathArrow	Tap, Hold, Swipe		x				x									
	Swipe	x					x									
DoubleEmphasisArrow	All						x						x			
BrokenArrow	All			x		x										
ArrowTip	All					x										
Chevrons	All					x						x		x	x	
ArrowTipWithCircle	All					x				x						
Chronophotography	All				x				x					x		
HighlightedZone	All								x						x	
HighlightedBorder	All								x				x			
DashedLines	Tap, Hold	x								x		x				
	Swipe, Flex											x				
DashedBalls	All									x		x		x		
CursorWithSupport	Tap, Hold									x	x					
	Swipe, Flex						x				x					
CursorWithoutSupport	Tap, Swipe, Flex										x					
	Hold									x	x					
MovementStroke	Tap, Swipe, Hold	x														
	Flex	x										x				
EmphasisLines	Tap, Hold									x						
	Swipe, Flex							x								
Ring	All								x	x					x	
MatchingShapes	All									x					x	
ElectricCharges	All									x	x				x	x
Gradient	All	x											x			
Target	Tap									x	x				x	
	Swipe, Flex, Hold					x					x				x	

Table 5. Categories matched by representations according to the considered microgesture and family. "All" designates `tap`, `hold`, `swipe` and `flex` together.

B ONLINE EXPERIMENT ANALYSIS : PARTICIPANTS' FAMILIARITY (AUGMENTED REALITY/ MICROGESTURES) AND IMPACT OF THE ORDER OF PRESENTATION (MICROGESTURES/ STATES)

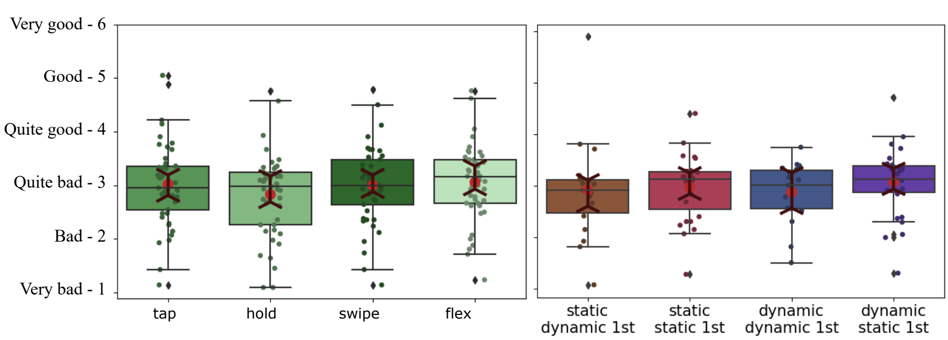
The online experiment took into account participants' familiarity with the topics, i.e. AR and microgestures. We clustered the answers we had in two groups and did not observe any significant impact of this familiarity as depicted by Figure 13. Figures 14, 15a and 15b show that neither the order in which microgestures were presented, nor the order of the representation state, i.e. *static 1st* or *dynamic 1st*, have a significant impact on the results.

Figure 15c brings proof that no microgesture global score was impacted by a specific representation state, i.e. *static* or *dynamic*. Figure 16 provides more detail on the mean and standard deviation for each family. The results are illustrated by clusters made of the family representations among the microgestures and presented in different plots according to the state and state order. We chose to use the same format than in the paper for reporting the values but preferred bar plots to box plots for Figure 15 to better show the grouping of values.



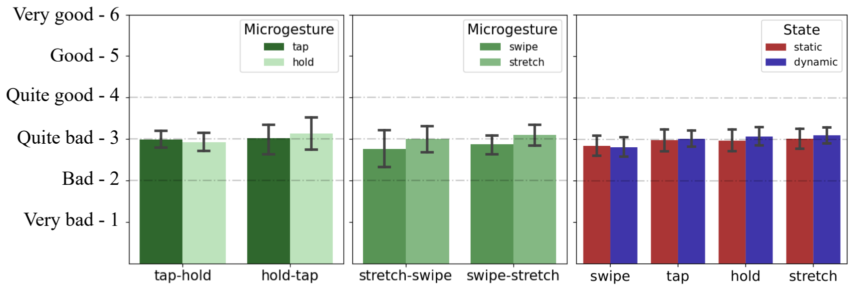
(a) Familiarity for AR (b) Familiarity for microgestures

Fig. 13. Distributions of the participants' mean scores per familiarity.



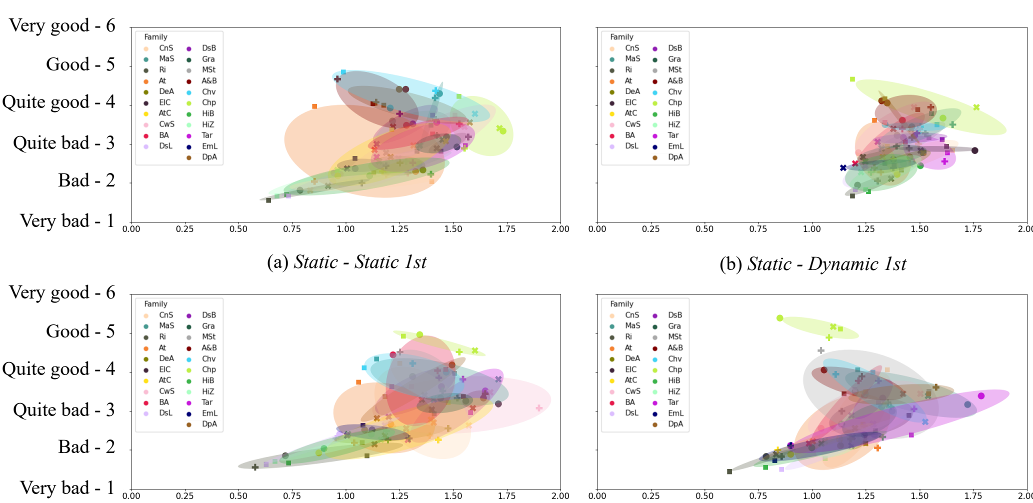
(a) Distributions per microgesture (b) Distributions per state and state order

Fig. 14. Distributions of the participants' mean scores.



(a) Per presented order for tap and hold (b) Per presented order for swipe and stretch (c) Per state and microgesture

Fig. 15. Mean reported quality scores of the representations and their 95% confident intervals.



(a) *Static - Static 1st* (b) *Static - Dynamic 1st*

(c) *Dynamic - Static 1st* (b) *Dynamic - Dynamic 1st*

Fig. 16. Bivariate density ellipses with a 50% coverage for family scores.
The x-axis represents the standard deviation and the y-axis the mean.

C FAMILY EXPLICITNESS

Each family uses different strategies to convey information about the microgesture motion as introduced in the paper. Tables 6, 7, 8 and 9 detail the calculation of the explicitness for each microgesture representation, ordered by family.

Family	*A&B*	*DpA*	*DeA*	*BA*	*At*	*Chv*	*AtC*	*Chp*	*HiZ*	*HiB*	*DsL*
Explicitness	0,67	1	0,5	1	0,33	0,67	0,33	0,67	0,17	0,17	0,17
Trajectory	x	x		x		x		x			
Direction	x	x		x		x					
Actuator	x	x		x		x		x			
Receiver	x	x	x	x	x	x	x	x	x	x	x
Release		x	x	x	x		x	x			
Separate		x	x	x							

Family	*DsB*	*CwS*	*CnS*	*MSt*	*EmL*	*Ri*	*MaS*	*ElC*	*Gra*	*Tar*
Explicitness	0,67	0,33	0,33	0,67	0,17	0,17	0,33	0,5	0,5	0,5
Trajectory	x			x					x	
Direction										
Actuator	x			x			x	x	x	x
Receiver	x	x	x	x	x	x	x	x	x	x
Release	x	x	x	x				x		x
Separate										

Table 6. Decomposition of explicitness by family for the tap microgesture.

Family	*A&B*	*DpA*	*DeA*	*BA*	*At*	*Chv*	*AtC*	*Chp*	*HiZ*	*HiB*	*DsL*
Explicitness	1	0,83	0,33	1	0,33	1	0,33	0,67	0,17	0,17	0,33
Trajectory	x	x		x		x		x			
Direction	x	x		x		x					
Actuator	x	x		x		x		x			
Receiver	x	x	x	x	x	x	x	x	x	x	x
Pause	x	x	x	x	x	x	x	x			x
Separate	x			x		x					

Family	*DsB*	*CwS*	*CnS*	*MSt*	*EmL*	*Ri*	*MaS*	*ElC*	*Gra*	*Tar*
Explicitness	0,67	0,5	0,5	0,67	0,33	0,5	0,67	0,67	0,83	1
Trajectory	x			x					x	x
Direction										x
Actuator	x			x			x	x	x	x
Receiver	x	x	x	x	x	x	x	x	x	x
Pause	x	x	x	x	x	x	x	x	x	x
Separate		x	x			x	x	x	x	x

Table 7. Decomposition of explicitness by family for the hold microgesture.

Family	*A&B*	*DpA*	*DeA*	*BA*	*At*	*Chv*	*AtC*	*Chp*	*HiZ*	*HiB*	*DsL*
Explicitness	1	0,67	0,33	0,67	0,33	0,67	0,67	0,33	0,33	0,33	0,33
Trajectory	x	x		x		x		x	x	x	x
Direction	x	x	x	x	x	x	x				
Explicit touch point	x						x				

Family	*DsB*	*CwS*	*CnS*	*MSt*	*EmL*	*Ri*	*MaS*	*ElC*	*Gra*	*Tar*
Explicitness	0,67	1	0,33	0,33	0,33	0,33	0,67	0,67	0,33	1
Trajectory	x	x		x			x	x	x	x
Direction		x								x
Explicit touch point	x	x	x		x	x	x	x		x

Table 8. Decomposition of explicitness by family for the `swipe` microgesture.

Family	*A&B*	*DpA*	*DeA*	*BA*	*At*	*Chv*	*AtC*	*Chp*	*HiZ*	*HiB*	*DsL*
Explicitness	1	0,67	0,67	0,67	0,33	0,67	0,67	0,33	0,33	0,33	0,33
Trajectory	x	x		x		x		x			
Direction	x	x	x	x	x	x	x				
Explicit moving part	x		x				x		x	x	x

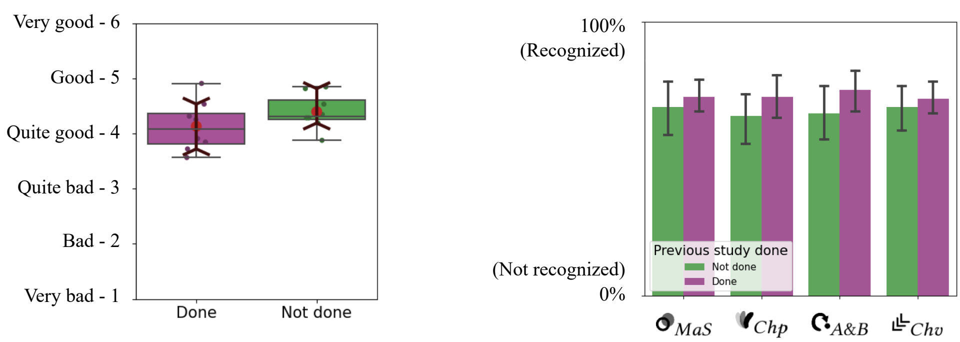
Family	*DsB*	*CwS*	*CnS*	*MSt*	*EmL*	*Ri*	*MaS*	*ElC*	*Gra*	*Tar*
Explicitness	0,33	1	0,67	0,33	0,67	0,33	0,67	0,67	0,33	1
Trajectory	x	x		x			x		x	x
Direction		x	x		x			x		x
Explicit moving part		x	x		x	x	x	x		x

Table 9. Decomposition of explicitness by family for the `flex` microgesture.

D HEADSET EXPERIMENT ANALYSIS : PARTICIPANTS' PRIOR KNOWLEDGE ABOUT MICROGESTURE REPRESENTATIONS

The headset experiment took into account that half of the participants had participated to the online study. As shown in Figure 17, we did not observe a significant impact of this prior knowledge on the microgesture representations.

We chose to use the same format than in the paper for reporting the values but preferred bar plots to box plots for Figure 17b to better show the grouping of values.



(a) Distribution of the scores given to each representation

(b) Distribution of the recognition rates per family

Fig. 17. Distribution of the participants mean scores depending on whether they had participated (Done) or not (Not done) to the online experiment.