Computers in Human Behavior 140 (2023) 107588

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Touchable pixels: Examining the affordance effect between an on-screen object and a user-elicited gesture on the touchscreen

Sangyeon Kim [a], Sangwon Lee [b,*]

[a] Department of Interaction Science, Sungkyunkwan University, Seoul, South Korea

[b] School of Industrial and Management Engineering, Korea University, Seoul, South Korea

ARTICLE INFO

Keywords:
Affordance
Touch gesture
Digital materiality
Gesture elicitation
Correspondence effect

ABSTRACT

Touching a smartphone screen is one of the most frequent activities in our everyday life. These touch gestures have recently been investigated under the user-centered approach, which collects gestures proposed by end-users. Despite its advantages in cognitive and behavioral performance, these gestural studies neglect the effect of visual properties of interface objects, i.e., digital affordances. This study aims to explore the effect of digital affordances on the screen through two sequential studies. In the preliminary study, an online questionnaire and interview were conducted to investigate user-defined gestures. Twenty participants responded to the survey to select the most appropriate gesture in the combination between four functions and five gestures for eight stimuli manipulated in several visual properties. The collected gestures for each function were statistically compared in terms of visual properties. In the main study, a behavioral experiment was conducted to examine the object-based correspondence effect. Twenty-six participants executed real touch gestures on the eight stimuli, and the gestural response time was collected to compare corresponding and non-corresponding conditions between gestures and stimuli. The results of both the preliminary and main studies indicated significant effects of visual properties in gesture execution. We concluded that digital affordance exists on the touchscreen; furthermore, it has unique characteristics grounded on the hybrid materiality of a digital environment consisting of both hardware and software. The new findings on digital affordance can be summarized as 1) a digital entity is perceived as freely manipulable regardless of physical laws, 2) a visuo-perceptible entity can be only perceptually constrained by other visual obstacles, and 3) the gestural priority depends on the embodied direction inherent in particular content. This study contributes theoretically to a better understanding of digital affordance and practically designing touch gestures based on the characteristics of affordances on the screen.

1. Introduction

Touch gestures are the most ubiquitous way to interact with smartphones, tablets, kiosks, and other screen-based interfaces. These interface gestures enable users to execute various specific functions depending on their intentions. Thus, designing these gestures needs to be studied in the context of users' behavioral and psychological characteristics. Yet, early interface gestures were pre-defined by system designers without any consideration of users, and have been criticized for their arbitrary gesture mapping, which neglects users' mental models (Morris, Wobbrock, & Wilson, 2010). In response to this criticism, a user-centered approach was introduced. This approach collects various gestures elicited by inexperienced or less experienced users on a specific technology and defines a set of the gestures in accordance with the degree of user agreement. This gesture elicitation method has shown good performance, high user preference, and other cognitive advantages such as memorability (Morris et al., 2010; Williams & Ortega, 2020; Wobbrock, Morris, & Wilson, 2009), thus being a major research paradigm for designing interface gestures.

Most gesture elicitation studies have tended to merely focus on the relation between a function and a user-defined gesture, neglecting the effect of perceptual properties of an interface object—affordance (Kim & Lee, 2020b). Specifically, a few researchers have introduced gestural effects of physical affordances in tangible media (Sharma, Roo, & Steimle, 2019; Soni et al., 2020); but researchers have rarely investigated the on-screen affordance in touchscreen devices. However, the visual properties of an on-screen object seem to be influential in gesture execution. For example, a hyperlink is easily distinguished from other

* Corresponding author. #417 Innovation Hall, 145 Anam-ro, Seongbuk-gu, Seoul 02841, South Korea.
E-mail address: upcircle@korea.ac.kr (S. Lee).

https://doi.org/10.1016/j.chb.2022.107588
Received 30 March 2021; Received in revised form 16 November 2022; Accepted 28 November 2022
Available online 30 November 2022

plain texts despite small differences in visual properties, an underline and blue color. These visual properties guide users to move to another page through a simple tap gesture. This implies that an object's visual properties themselves play a determinant of perception on its function and operation. Moreover, this effect of visual properties was recently observed in one gesture elicitation study (Kim & Lee, 2020b). This study found that some visual properties, layout and number of entities, significantly influenced a user's gesture elicitation. In this context, the user-defined gestures should be investigated in accordance with the effect of on-screen affordances.

This article aims to examine the affordance effect on the manipulation of a touchscreen interface. For this purpose, we preliminarily investigate user-elicited gestures between four functions and five gestures on eight stimuli manipulated in several visual properties. Based on the results of the preliminary study, we further conduct a behavioral experiment to examine an object-based correspondence effect. The object-based correspondence effect refers to the effect that potentiates corresponding actions with an object's perceptible properties at the perceptuomotor level. Specifically, this correspondence effect allows subjects faster and more accurate responses in a corresponding condition between an object and an action, than a non-corresponding condition. For instance, several corresponding effects have been reported, such as the alignment between a handle grip orientation and the left or right hand (Azaad & Laham, 2019; Bub & Masson, 2010; Tucker & Ellis, 1998), volumetric compatibility between object size and grip type (Tucker & Ellis, 2001; Vainio, Symes, Ellis, Tucker, & Ottoboni, 2008), and other correspondences between an object's properties and a subject's actions. Thus, the main study investigates a corresponding effect between a visual property of interface objects and a touch gesture on the screen. For this experiment, the corresponding and non-corresponding gestures are separated based on the most and least agreement toward a particular stimulus in the preliminary study. We expect that the corresponding effect will be observed if there is an effect of digital affordance on the screen.

2. Background

2.1. Digital affordance

People interact with an environment and discover value or meaning, actualizing their intention on the environment through their actions. To explicate this phenomenon, Gibson (1979) coined the term *affordance*, which refers to an action possibility providing a particular value or meaning innate in an environment for a perceiver. In other words, affordance is a combination of physical properties (e.g., size, texture, weight) affording a perceiver valuable and meaningful action. For instance, a rock, which has a flat surface and solid substance and appropriate height to be a seat for a human, offers value for a perceiver by functioning as a chair, bench, or other seat. These physical material properties of the rock determine how the perceiver can act on the object, providing sit-ability, lie-ability, and other action possibilities. Thus, affordance can be defined as a value-providing action possibility based on perceptual patterns inherent in an environment.

The concept of affordance was first introduced into the interaction design and human-computer interaction (HCI) fields by Norman (1988). This concept has fascinated researchers and practitioners in terms of designing a product providing a natural feedforward for increasing intuitiveness (Hornecker, 2012; Hurtienne & Blessing, 2007; Still, Still, & Grgic, 2015; Vermeulen, Luyten, van den Hoven, & Coninx, 2013). Specifically, designing with this concept enables a user to achieve a goal to easily interact with a product following a designer's intention, thereby increasing usability and usefulness (McGrenere & Ho, 2000). In this context, researchers have introduced new designing methods based on the concept of affordance for technological artifacts. These design methods aim to develop a physical artifact with improved ease of use based on the understanding of relations between a product's mechanical parts or technical functions and a user's tasks (Galvao & Sato, 2005; Maier & Fadel, 2003; Masoudi, Fadel, Pagano, & Elena, 2019).

Despite these theoretical and practical implications of the design of a physical artifact, the concept of affordance has been considered theoretically incompatible with designing screen-based interfaces, deriving an ontological controversy on the digital affordance. For instance, Norman (2013) has criticized that the concept of affordance has been widely misused by interface designers who utilized this concept on the digital object on the screen. He argued that an object depicted on the screen cannot provide affordances because of the lack of physical constraints in the object image, devaluing the interface object as a mere sign (Norman, 2008). According to him, the only affordances on the screen can be derived from the material properties of the glass panel, which can be touched anywhere on the screen, regardless of an interface object. Instead of this physical affordance, he suggested that the manipulation of digital objects on the screen mostly depends on the long-term memory in a user's past experience interacting with digital signs, denying the concept of digital affordance.

However, this argument neglects the integrated material properties of both hardware and software in interactive media (Jung, Wiltse, Wiberg, & Stolterman, 2017). Most of the software and its interface objects provide numerous action possibilities and constraints pre-defined by system designers or its manufacturers. A button is programmed to be tapped in a system when any valid input signal is detected in that position on the screen. This programmed rule can be discovered by users exploring the digital environment with their fingers. In the repetitive uses of the button, users can learn to perceive the affordance tap-ability that looks like a button with a rectangular shape and appropriate size for a finger. Yet, this kind of affordance is determined by software as well as hardware. If there is no capacitive touch panel to receive a static electricity signal from a user's finger in the screen, users never discover the tap-ability and other touch-ability of the button. This shows that an affordance in digital technology cannot provide new affordances beyond their hardware capabilities. Thus, it is reasonable that the digital affordance be considered as the collaborative material properties of not only hardware, but software as well. In this vein, affordance in digital media should be understood based on complex technical characteristics to support user's activities (Allen-Robertson, 2018; Berriman & Mascheroni, 2019).

2.2. Object-based correspondence effect for an interface object

The object-based correspondence effect refers to the effect derived from an object's innate properties to invoke faster and more accurate action responses on corresponding stimuli than non-corresponding stimuli (Kim & Lee, 2020a; Proctor, Lien, & Thompson, 2017; Tucker & Ellis, 1998). This correspondence effect occurs in motor tasks interacting with an object, thus influencing an agent's action. More specifically, the agent's action is determined by the compatibility between the object's physical properties and the agent's action capabilities. If there is a compatibility between the object's features and the agent's action, the agent can naturally act on the object, showing a faster response; whereas, if there is an incompatibility, the agent is interrupted to act on the object, showing a slower response. It demonstrates that an object's physical properties intuitively evoke a motor representation of an agent if the object has an appropriate structural feature to be manipulated in a specific way.

For example, some studies have reported that there is an effect of alignment between a handle grip orientation and the use of the left or right hand (Azaad & Laham, 2019; Bub & Masson, 2010; Tucker & Ellis, 1998). In these studies, participants responded faster and more accurately when they use a left hand on the left-oriented handle of an object, whereas a right hand on the right-oriented handle of an object. Other studies have reported that there is an effect of volumetric compatibility between object size and grip type (Tucker & Ellis, 2001; Vainio et al., 2008). In these studies, participants responded faster and more accurately on the thick object with largely grasping action using a whole hand, whereas the thin object with pinch action using mere two fingers.

These object-based correspondence effects have been attributed to automatic perceptuomotor processing triggered by an object's properties (e.g., size, orientation, shape)—affordance. However, this interpretation has been criticized in terms of the experimental paradigm using a keypress task. Some researchers have argued that the keypress task cannot evoke a motor representation of grasp-ability on a stimulus because of the qualitative difference between pressing a key on the button box and a grasping action. They claimed that these effects are rather derived from a spatial coding based on visual saliency or asymmetry, not from an affordance innate in a stimulus (Cho & Proctor, 2010; Proctor & Miles, 2014). Thus, a reach-and-grasp task was introduced for the experiment to examine the object-based correspondence effect (Bub, Masson, & Kumar, 2018, 2021; Chong & Proctor, 2020). Experiments with a reach-and-grasp task make participants execute an actual action to reach and grasp toward a stimulus. In a comparative study, this reach-and-grasp task was attributed to successfully evoke the correspondence effect by the participants' intention to act on a stimulus, whereas the button-press task did so by perceptual saliency (Pavese & Buxbaum, 2002). Thus, it indicates that a behavioral context significantly contributes to evoke a genuine motor affordance on a stimulus (Bub, Masson, & van Noordenne, 2021; Chong & Proctor, 2020).

This study aims to examine an object-based correspondence effect of an interface object on a touchscreen. Although the correspondence effect has rarely been studied for an interface object compared to a physical object, this effect might be observed if there is an affordance to touch an on-screen object in a particular way. To examine this effect, the experiment should be conducted under the exactly separated corresponding and non-corresponding conditions between an interface object and a touch gesture. Yet, this separation is somewhat problematic because there is an ambiguity in whether or not people can manipulate an on-screen object through a particular gesture. This ambiguity is derived from the characteristic of a digital affordance grounded on both the characteristics of the hardware and the software. In respect to the hardware, users can make any touch gestures on the screen; in respect to the software, there are subtle different gestural rules set by each system designer or each manufacturer. This difference implies that there is no criterion for separating experimental conditions. To establish this criterion, we suggest the preliminary study on user-elicited gestures on stimuli manipulated in some visual properties. This preliminary study will allow us to explore object-gesture mappings in an interface, thus separating corresponding and non-corresponding conditions based on the users' agreed mappings between touch gestures and interface designs.

3. Research questions

Given the background above, the concept of affordance can be valid in a digital environment that has its own material characteristics and rules. However, the digital material on the screen has not been empirically explored, but rather is often neglected. Therefore, this present study attempts to establish whether an on-screen affordance exists or not, and furthermore, to explore significant behavioral effects of visual properties via examining an object-based corresponding effect. Based on these goals, we propose two research questions as follows:

Research Question 1: Does an on-screen object afford particular gestures for a specific function?

Research Question 2: How do particular visual properties affect on-screen object manipulation at the perceptuomotor level?

To answer the research questions, this study attempts to examine the gestural effect of digital affordances on the touchscreen systemically in two steps. First, in the preliminary study, we exploratively elicit representative mappings between functions and gestures in terms of visual patterns. If there is an effect of visual properties, different screen patterns might cause different gestures for a particular function. Next, in the main study, we conduct a behavioral experiment comprising comparative conditions based on the preliminary study results. This experiment might provide useful insight into digital affordances at the perceptual level of individual visual properties through examining an object-based correspondence effect.

4. Preliminary study

4.1. Overview

This preliminary study aims to approximately understand users' mental models on function-gesture mappings in terms of visual patterns. For this purpose, we conduct an exploratory sequential mixed method research consisting of quantitative and qualitative methods. First, an online survey was first conducted to collect and analyze these function-gesture mappings. Specifically, this survey constituted several questions asking participants to match appropriate gestures for functions on stimuli manipulated in several visual properties, referring to gesture elicitation studies. Next, a follow-up interview was conducted to enhance the interpretation of the quantitative results. Both results of this study are expected to provide valuable insights on the potential effect of visual properties. Moreover, the results on function-gesture mappings will be applicable to the design for the experimental conditions in the subsequent study on the object-based correspondence effect. Based on the results of this study, the most agreed upon gestures on a particular function will be categorized as corresponding conditions, whereas the worst or less agreed upon gestures will be categorized as non-corresponding conditions in the main study.

4.2. Method

4.2.1. Participants

Twenty graduate students (15 females, 5 males) were recruited in Sungkyunkwan University, South Korea. The students were voluntarily involved in this study through a notice on bulletin boards in the university. All the students who contacted us for participation received basic instruction and conducted the online survey. The participants ranged in age from 24 to 38 (M = 27.5, SD = 3.1). They have interdisciplinary backgrounds among computer science, social science, interaction design, and artificial intelligence. None of the participants have any visuomotor impairment in the interaction with digital devices. They were compensated with 3000 KRW for participation.

4.2.2. Data collection

To collect representative mappings between gestures and functions in terms of visual properties, we followed the gesture elicitation paradigm to ask end-users to select an appropriate gesture on a function (Morris et al., 2010; Wobbrock et al., 2009). This gesture collection in this study was conducted through an online survey that was created in Google Forms. The questionnaire consisted of totally 32 questions for function-gesture mappings on eight stimuli manipulated in visual properties. Specifically, one stimulus has four questions corresponding to the number of functions—view, back, remove, and next. Each question on one function was "which gesture do you prefer for this screen when you execute this function?" with five-choice options—tap, double tap, long press, slide, and scroll. To sum up, a participant was asked to select one preferred gesture for one stimulus to execute a particular function in each question. Thus, we eventually collected 32 responses (eight stimuli on the four function-gesture mappings) from each participant, totally 640 responses from the twenty participants. All the stimuli were presented in a random order to avoid an order effect. In addition, to minimize the nuisance effect of stimulus size, the participants were instructed to conduct the survey in a laptop or desktop environment with full HD resolution. In these environments, the stimulus was presented in size similar to most existing smartphone screen sizes.

After the online survey, the participants were invited to attend a follow-up interview in person. The interview schedule was individually negotiated between participants and the interviewer. The one-to-one interviews were conducted in a private room at our laboratory in Sungkyunkwan University. The interviews ranged between 20 min and 30 min in length. The semi-structured interview was conducted to enhance understanding of the results of the quantitative data analysis. At the beginning of an interview, a participant was first informed about the goal of this interview and then asked to write consent of interview participation. In the main session, the participants were asked a series of close- and open-ended questions focusing on 1) their intentions to select a particular gesture on each function in the survey, and 2) their actual experiences interacting with touchscreen devices on the function-gesture mapping. More specifically, the researcher first asked, *"what kind of gesture did you choose for this function on this screen?"* to confirm their responses of interest. After then, the researcher asked the reason, *"why did you choose the gesture? Could you explain it in detail?"* Lastly, the researchers asked to share their experience, *"If you can remember any experience of this function-gesture mapping, could you tell your real story?"* All these interviews were digitally audio-recorded and transcribed into text. In the transcript, personal information of all participants was anonymized or removed for their privacy.

4.2.3. Stimulus material

The stimuli consist of images of a smart device with a square-shaped screen, and on-screen design patterns were manipulated in visual

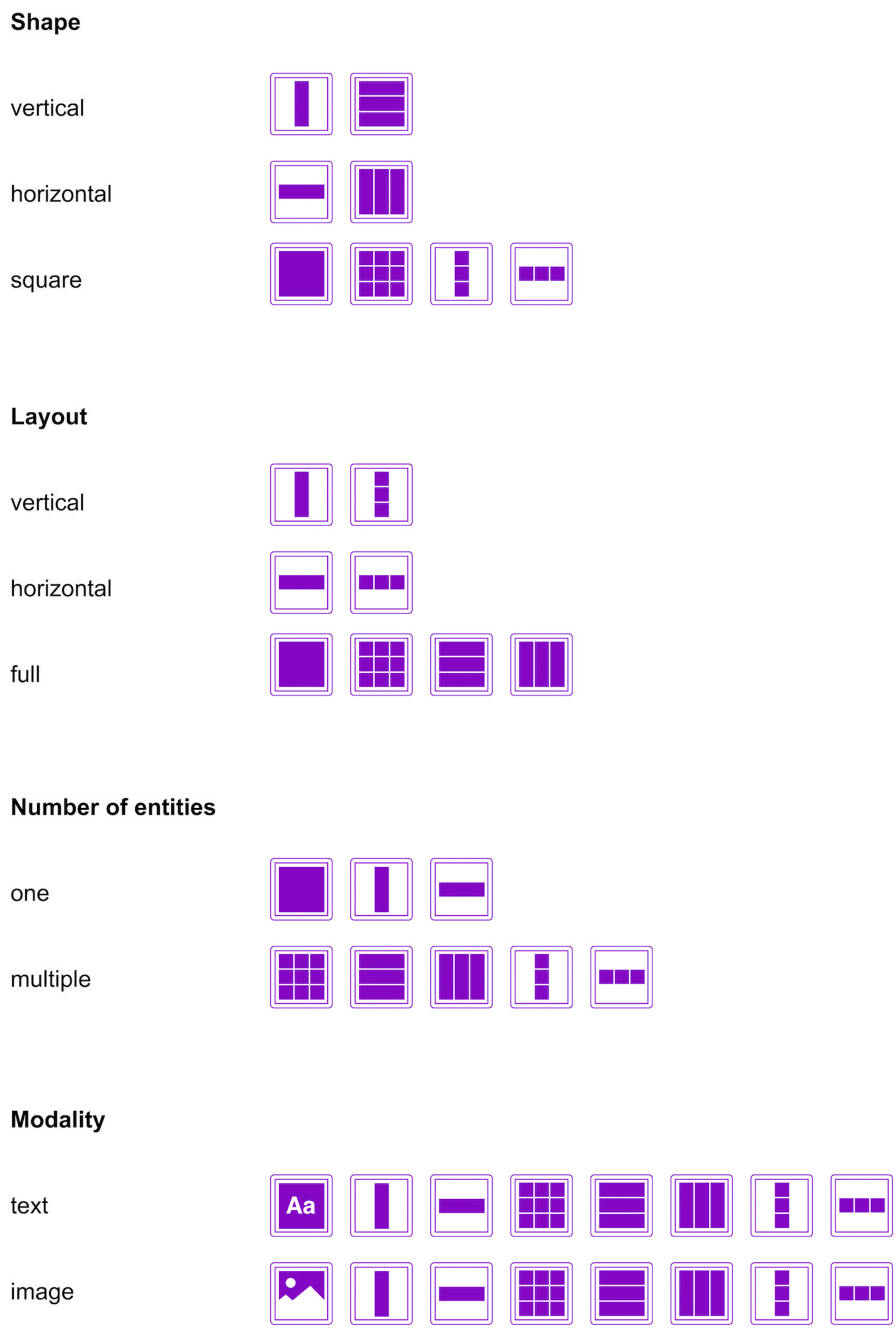

Fig. 1. Stimulus manipulation in terms of four visual properties: 1) shape (vertical, horizontal, and square), 2) layout (vertical, horizontal, and full), 3) number of entities (one and multiple), and 4) modality (text and image).

properties such as shape, layout, and number of entities. The unusual shape of the device with a square-shaped screen was intended to control the effect of screen asymmetry between width and height. Despite numerous combinations with those properties, we eliminated impossible and unrealistic combinations in consideration of the real-world interface design, thus conclusively determining eight stimuli. The sets of stimuli for each visual property are fully arranged in Fig. 1.

To strictly compare the effect of interface modality, the stimuli were created in two types of interface modality, image and text, and the participants were separately assigned to the questionnaire on the stimuli on only one modality. This separation was intended based on the literature on the effect of modality for a user's perception and interaction (Kim, Sundar, & Park, 2011; Sundar, Xu, & Bellur, 2010). Thus, our stimuli for the text modality consist of cards that include title, date, and a part of body text. To reduce the potential effect of a word or sentence perception, all texts in the card were filled with meaningless texts, *lorem ipsum*, which is commonly used as a filler for text fields in graphic design. The stimuli for the image modality consist of several photographs of natural and urban landscapes all over the world. The design patterns of all the stimuli were created with reference to some applications that exhibit one or multiple content such as an image gallery, a newsfeed on social media, or message applications. The stimuli are shown in Fig. 2.

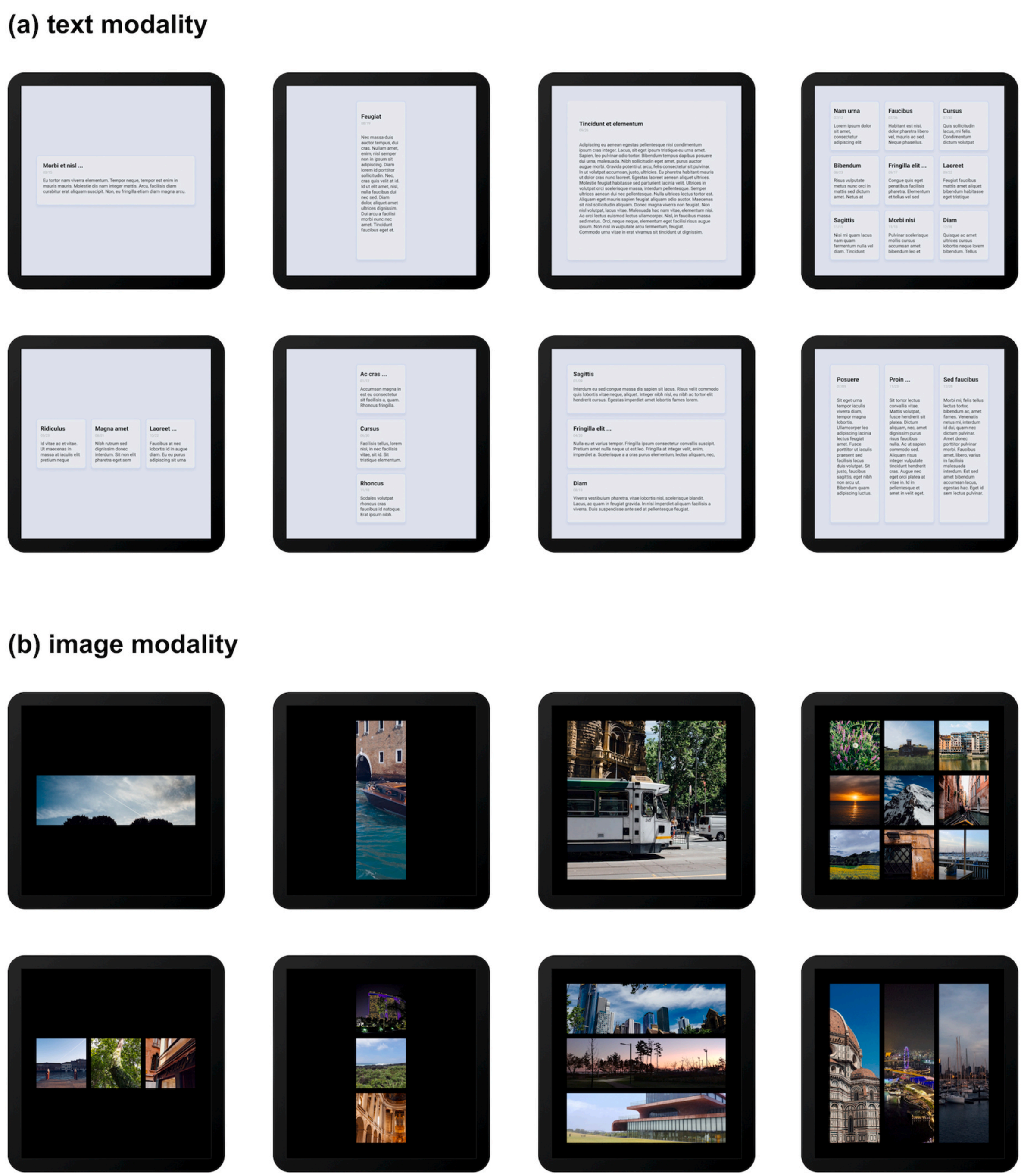

Fig. 2. The entire list of sixteen stimuli used in the survey. The same eight design patterns were applied for (a) text modality with random filler texts and (b) image modality with random landscape photographs.

4.3. Results

4.3.1. Overall gesture-function agreement

This section aims to explore the overall tendency of function-gesture mappings on total stimuli. For this purpose, two analyses were conducted: agreement rate (AR) analysis and corresponding analysis. The AR analysis is the most prevalent method to quantify the degree of gesture agreements on functions. Although there are some statistical methods (e.g., Fleiss kappa) to analyze the degree of agreements for multiple raters, gesture elicitation research tends to use the agreement indices proposed by Wobbrock and Vatavu and their colleagues (Vatavu & Wobbrock, 2015; Wobbrock et al., 2009). The most recently refined formula is presented as follows:

$$AR(f) = \frac{|G|}{|G|-1}\sum_{G_i \subseteq G}\left(\frac{|G_i|}{|G|}\right)^2 - \frac{1}{|G|-1}$$

In this formula, G refers to the set of overall gestures selected by all participants on each function f. $|G|$ is the size of set G, and $|G_i|$ is the sizes of subsets of identically selected gestures. On the basis of this AR formula, we compute the AR score for each function in the online survey and set the scores out in Table 1. Vatavu and Wobbrock (2015) categorized classes for the AR score based on the probability distribution of AR; specifically, AR < 0.100 as low, $0.100 <$ AR < 0.300 as medium, $0.300 <$ AR < 0.500 as high, and AR > 0.500 as very high.

In addition to the ARs, we further conducted corresponding analysis to understand the complicated relations between four functions and five gestures. The results are presented in Fig. 3. The graph shows mappings for total responses of all stimuli on gestures for a function. The closer a function and gesture on the graph, the more associated the gesture and function. Points that are farther away from each other indicate that they are perceived in different type of gestures or functions.

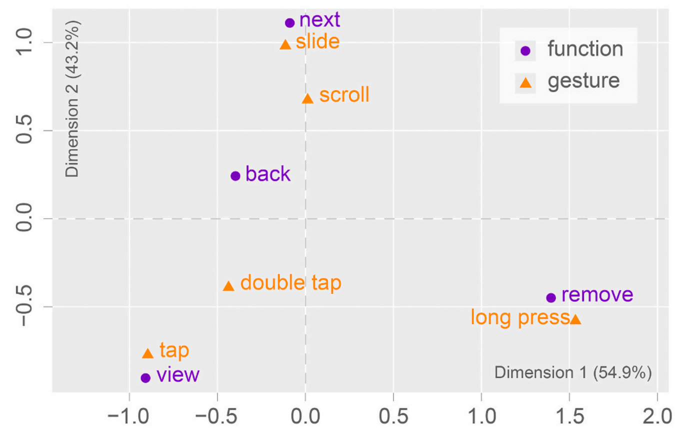

Fig. 3. Results of the corresponding analysis between functions and gestures. It shows how much functions and gestures are overall associated in participants' perceptions: the closer a function and a gesture, the more associated the pair.

4.3.2. Effects of visual properties on each function

This section investigates the effect of visual properties on each function. A Chi-square test was primarily conducted for the comparison among gestures with each visual characteristic, shape, layout, number of entities, and modality. For more reliable analysis, we eliminated rarely selected gestures as outliers, which have an expected frequency of less than five in common. Despite this preprocessing, if there was a violation of the assumption for the Chi-square test, a Fisher's exact test or Freeman-Halton extension of Fisher's exact test was complementarily conducted.

View. The majority of the selection for 'view' was tap (86.6%), and some of them were double tap (10%), and the others were long press (1.9%), scroll (2.5%), and nobody selected slide. The rarely selected gestures, long press, scroll, and slide, were eliminated before the analyses. A Chi-square test revealed a significant difference in the gesture selection between one entity and multiple entities ($\chi^2_{(1,N=153)} = 11.4, p < 0.001$). Double tap is more frequently selected on the stimuli with one entity rather than multiple entities. In addition, a Chi-square test revealed a significant difference in the gesture selection between text and image modalities ($\chi^2_{(1,N=153)} = 5.33, p = 0.021$). Double tap is more frequently observed on the stimuli with image modality than those with text modality. To synthetically consider these two significant effects, Fisher's exact tests were conducted. In the results, only the image modality showed a significant difference in the gesture selection between one entity and multiple entities (Fisher's exact test, $p < 0.001$), whereas the text modality did not (Fisher's exact test, $p < 0.639$). Participants selected double tap more frequently on the stimuli with one image than other conditions (Odds ratio = 15.33, 95% Confidence interval = 3.02–78). Lastly, there was no significant effect in both shapes (Freeman-Halton extension of Fisher's exact test, $p = 0.693$) and layout (Freeman-Halton extension of Fisher's exact test, $p = 0.693$).

Table 1
Results of the agreement rate analysis. Each function has the most agreed gesture with its agreement rate (AR) score. The AR score shows how much a pair between a function and a gesture is overall agreed by participants.

Function	Most Agreed Gesture	AR(i)
View	Tap	0.743
Back	Slide	0.264
Remove	Long Press	0.645
Next	Slide	0.527

Back. The most selected gesture for 'back' was slide (33.1%), the second was tap (30%), the next one was scroll (21.9%), then double tap (14.4%), and lastly long press (0.6%). The rarely selected gesture, long press, was eliminated before the analyses. In the results, there was no systematic effect of visual properties, shape ($\chi^2_{(3,N=159)} = 2.98, p = 0.811$), layout ($\chi^2_{(3,N=159)} = 1.71, p = 0.944$), number of entities ($\chi^2_{(3,N=159)} = 1.69, p = 0.693$), and modality ($\chi^2_{(3,N=159)} = 1.5, p = 0.683$).

Remove. The majority of the selection for 'remove' was long press (79.4%), some of them were scroll (10.6%) and slide (6.8%), and the other was double tap (3.1%), and nobody selected tap. The rarely selected gestures, tap and double tap, were eliminated before the analyses. In the results, there was no significant effect of visual properties, shape (Freeman-Halton extension of Fisher's exact test, $p = 0.663$), layout (Freeman-Halton extension of Fisher's exact test, $p = 0.253$), number of entities ($\chi^2_{(2,N=155)} = 2.73, p = 0.255$), and modality ($\chi^2_{(2,N=155)} = 2.04, p = 0.361$).

Next. The majority of the selection for 'next' was slide (63.1%), some of them were scroll (36.3%), then tap (0.6%), and nobody selected double tap and long press. The rarely selected gestures, tap, double tap, and long press, were eliminated before the analyses. In the results, we found significant effects for shape and layout. A Chi-square test revealed a significant difference in the gesture selection among the square and horizontal and vertical shapes of each entity in a stimulus ($\chi^2_{(2,N=159)} = 9.12, p = 0.010$). Specifically, slide is more frequently selected on the stimuli with vertical shape entities, whereas scroll is more frequently selected on the stimuli with horizontal shape entities. In addition, a Chi-square test revealed a significant difference in the gesture selection among full, horizontal, and vertical layouts of entire entities in a stimulus ($\chi^2_{(2,N=159)} = 12.7, p = 0.002$). Slide is more frequently selected on the stimuli with a horizontal layout, whereas scroll is more frequently selected on the stimuli with a vertical layout.

To synthetically consider these two significant effects, Chi-square tests were conducted. Regarding the stimuli with a horizontal shape entity, a Chi-square test revealed a significant difference in the gesture selection between full and horizontal layouts ($\chi^2_{(1,N=39)} = 7.39, p = 0.007$). Participants selected slide more frequently on the stimuli with a horizontal layout, whereas scroll was more frequently selected on the stimuli with a full layout (Odds ratio = 6.50, 95% Confidence interval = 1.60–26.4). Regarding the stimuli with a vertical shape entity, a Chi-square test revealed a significant difference in the gesture selection between full and vertical layouts (Fisher's exact test ,$p = 0.044$). Participants selected slide more frequently on the stimuli with a full layout, whereas scroll was more frequently selected on the stimuli with a vertical layout (Odds ratio = 10.2, 95% Confidence interval = 1.12–93.3). Regarding the stimuli with a square shape entity, a Chi-square test revealed a significant difference in the gesture selection among full, horizontal, and vertical layouts ($\chi^2_{(2,N=79)} = 21.2, p < 0.001$). Participants selected gestures almost the same as the expected values on the stimuli with a full layout, but they selected slide more frequently on the stimuli with a horizontal layout, whereas scroll was most frequently selected on the stimuli with a vertical layout (Odds ratio = 91.2, 95% Confidence interval = 4.77–1754). Lastly, despite marginal significances at the 0.1 level, there was no significant effect in both numbers of entities ($\chi^2_{(1,N=159)} = 3.17, p = 0.075$) and modality ($\chi^2_{(1,N=159)} = 3.34$, $p = 0.068$).

4.4. Discussion

These results of the preliminary study indicated that there are significant effects of visual properties in the gesture selection for two functions 'view' and 'next.' More specifically, a gesture for 'view' was determined by number of entities and modality, and a gesture for 'next' was determined by shape and layout. In contrast, the functions 'back' and 'remove' did not show any significant effect of visual properties. We discuss these results for each function in detail.

The function 'view' shows the highest agreement according to the AR and the corresponding analysis. This function is very strongly associated with the gesture 'tap.' Despite the highest agreement, a significant number of participants selected double tap for the stimuli consisting of one image. We gained understanding of the results on the function 'view' through follow-up interview sessions. First, the gestural difference between modalities can be demonstrated by semantic ambiguity of the function. A participant who answered the questionnaire with the image modality mentioned that *"In my personal opinion, the image was expected to be enlarged (zoom in) and viewed the picture with more detail through double tap"* (P16), while another participant with the text modality mentioned that *"I chose tap to open the card and view entire texts in detail. (…) I never experienced to magnify a card through double tap. It is definitely impossible"* (P14). These imply that the function 'view' is separately interpreted as two meanings, opening and magnifying, in accordance with the modality. Thus, the word signifier 'view' seems to fail to anchor a particular significance of the function at the semantic level.

Second, the double tap on an image for the function 'view' might be derived from a legacy bias (Morris et al., 2014). A participant mentioned that *"Even though the stimuli were images of a smartphone or tablet screen, I chose double tap because the one image can be usually viewed (magnified) through double click in the PC environment. On the other hand, the multiple images reminded me of gallery apps on a smartphone. Because of that reason, I did not select double tap for the multiple images"* (P1). Another participant (P16) answered similarly to the above quote. This indicates that their responses were influenced by past experiences with several digital devices.

The function 'back' shows the least agreement compared to other functions. This function is not clearly associated with a specific gesture. This result is quite reasonable because users leave the current page and go back to the previous page on most mobile devices through a button 'back,' thus they are not expecting to use a specific gesture directly on the entire screen. Specifically, the button for 'back' is mostly represented as a less-than sign '<' or a left-directed arrow sign '←' sometimes with a text label and outline. When participants intended to execute 'back,' they might expect some targets such as icons, texts, or outlined buttons; but there was no such a target on the stimuli, thus deriving the least agreement.

In contrast to 'back,' the function 'remove' shows very high agreement. The function is strongly associated with the gesture 'long press' and weakly with 'slide' and 'scroll.' The highest association of 'long press' is understandable because this function is usually executed by a sequential task consisting of selecting one entity and removing the entity via a button. This gestural interaction can be observed everywhere in a photo gallery, note, message, and other mobile applications. Additionally, a few selections for 'slide' and 'scroll' were also plausible in terms of everyday use of smart devices. A push notification is commonly removed by a slide gesture and a running application is commonly closed by a scroll gesture. Yet, these function-gesture mappings are somewhat inconsistent in some apps such as Gmail and Apple Mail. The slide gesture in these apps provides different functions such as archiving messages. This kind of inconsistency seems to derive only a few selections for these gestures.

The function 'next' shows quite high agreement. This function is primarily associated with the gesture 'slide' and secondly associated with the gesture 'scroll'. The two gestures were differentiated in accordance with both visual properties, shape and layout. The shape of each entity invoked a specific gesture based on different layouts. Fig. 4 shows the overall mappings between each stimulus and the most agreed gesture. In these results, we observe a correspondence between the direction of the gesture and the visual flow of an entity or entities. These results indicate that participants might perceive the visual flow of the digital entities as one group, like a conveyer belt, to be manipulated together. This perceptual grouping can be explained that participants seemed to perceive continually arranged entities with a common fate to move together according to the Gestalt principles of continuity and common fate (Wagemans et al., 2012; Wertheimer, 1923). Yet, this idea does not explain why the 'slide' gesture was selected twice more than 'scroll' despite the same numbers of stimuli in the vertical and horizontal flows. We assumed that this result might be derived from the directions inherent in a photograph and a text. A photograph has several pieces of directional information such as horizontal or vertical lines in a landscape, and a text is written and read from left to right. This inherent direction might derive participants to have supremacy for 'slide' as compared to 'scroll' on the function 'next.'

Although the results indicate that there seems to be affordance

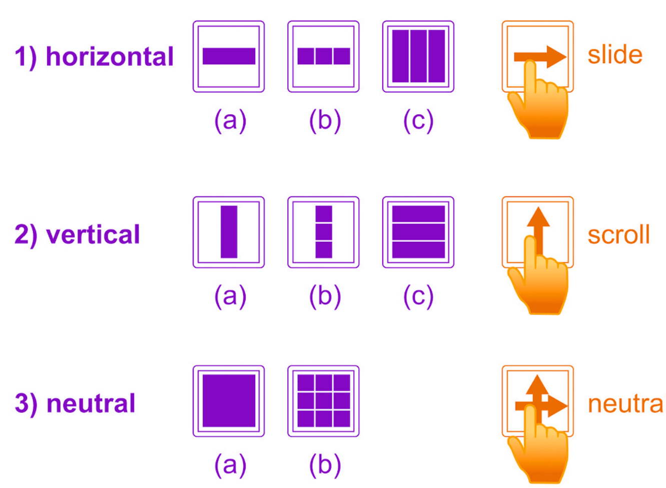

Fig. 4. According to the most agreed gesture for each stimulus, the newly suggested complex variable—visual flow consisting of three corresponding conditions between 1) horizontal and slide, 2) vertical stimuli and scroll, 3) neutral and any gesture.

effects on the screen, this study has a methodological limitation in which it is hard to strictly separate perceptuomotor effects from other cognitive mechanisms. This and other gesture elicitation studies have not controlled or limited a participant's response time. A participant was merely asked to suggest or select corresponding gestures with a particular function for enough time. However, this temporally unlimited condition might interrupt the genuine effect of affordance related to the perceptuomotor process of a human subject. Several psychologists have reported that perceiving affordance is behaviorally or neurologically processed under about 250 ms (Bub et al., 2018; Proverbio, Adorni, & D'Aniello, 2011). HCI scholars have similarly claimed that users feel an interface object to directly manipulate in 100 ms (Nielsen, 2009). This implies that our results in this preliminary study might be intervened by recall, inference, and other so-called higher cognitive mechanisms. For instance, the results of the function 'view' are partially derived from the legacy bias from past experiences with a desktop, specifically, explicit declarative memory. Yet, the concept of affordance is primarily established on the compatibility between an object's physical properties and a subject's bodily ability. Moreover, learning to perceive affordance has been considered to be related to implicit knowledge hard to be explicitly verbalized (Gibson, 2000; Kim & Lee, 2018; Still et al., 2015).

Therefore, we determined to further study an action-specific perceptual effect for the function 'next', which shows significant results but does not show any explicit consensus among participants in the follow-up interview. We expect that the following study will disclose more specific details of the gestural effect derived purely from the digital affordance regardless of other cognitive interruptions, thus focusing on the perceptuomotor effect of digital affordance inherent in visual patterns on a touchscreen.

5. Main study

5.1. Overview

This main study aims to precisely examine an object-based corresponding effect in terms of visual patterns. To investigate genuine motor affordance, a behavioral experiment was conducted to compare corresponding and non-corresponding gestures in response time for each visual stimulus (Bub et al., 2021; Chong & Proctor, 2020). These experimental conditions were determined in accordance with the results of the preliminary study. Specifically, the experiment was designed by three visual flows, which are complex visual patterns consisting of several visual properties and the appropriate gestures defined by end-users in Fig. 4. In this experiment, we used a modified version of the experimental tasks established by Bub et al. (2018). The overall procedure and tasks mostly originated from their experimental paradigm; but a word for the function 'next' was additionally presented at the initial trial once, and participants were asked to say the word repeatedly in every trial for priming participants' intentions to execute the function on stimuli. The experimental results are expected to provide details on the user's perceptuomotor characteristics on the interaction with digital objects manipulated in the visual flow and individual visual properties in detail.

5.2. Hypotheses

In the preliminary study, we found an interesting phenomenon that a visual flow of on-screen objects influences participants to select the gesture which has the same directional movement, although there is no physical constraint guiding to a specific directional gesture on the screen. On the relation between a visual flow and a corresponding gesture, we attributed this relation to the perceived direction of the visual continuum consisting of proximate entities under the Gestalt principles (Wagemans et al., 2012; Wertheimer, 1923). The perceptible properties guiding a direction innate in an interface object might allow users to perceive its movement possibility in this direction. This assumption seems quite reasonable because this direction-based correspondence effect has been continuously reported as various names such as control-display effect or directional stimulus-response effect in several control situations (Proctor & Vu, 2016). This directional effect generally occurs in behavioral compatibility between directions of control and display in the situation of controller manipulation. This effect enables an operator to manipulate a controller faster and more accurately when the control direction is congruent with the movement direction of the outcome than when it is incongruent.

This direction-based correspondence effect has been observed in several control situations. For example, in the machinery control such as an excavator or underground mining bolting machine, an operator manipulated a control level in a particular direction more naturally when the directions of the lever's movement and the machine's movement were compatible than an incompatible condition (Hoffmann & Chan, 2018; Steiner, Burgess-Limerick, & Porter, 2014). In addition to this machinery control, this effect has been found in virtual object controls. For example, a cursor control in a particular direction showed a faster and more accurate response when the joystick's movement direction was compatible than an incompatible condition (Worringham & Beringer, 1998). Moreover, this effect was also found in digital screen manipulation using a keyboard or a pen (Chen & Proctor, 2012, 2013; Janczyk, Xiong, & Proctor, 2019; Phillips, Triggs, & Meehan, 2005).

On these phenomena, researchers have claimed that this effect is fundamentally derived from the correspondence between movement directions of controlling limb and controlled elements in an operator's visual field (Hoffmann, Chan, Man, & Chan, 2019; Worringham & Beringer, 1998). In other words, this direction-based correspondence effect occurs when a controller has 1) perceptible information of its manipulability in a particular direction, and a display moves 2) compatibly with an operator's limbic movement. It seems that the touchscreen satisfies these conditions; specifically, it provides visual information to exhibit its changes of status to move in a direction immediately and supports direct manipulation of an interface object using a finger or fingers. It implies that the digital affordance of on-screen objects might be perceived by users, evoking a direction-based correspondence effect at the user's perceptuomotor level.

In this vein, according to the behavioral characteristic of the object-based correspondence effect, we can assume that the response time in the compatible condition of a direction between an on-screen pattern and a touch gesture will be faster than in the incompatible condition. More specifically, the response time will be significantly faster when a subject executes a gesture that has a specific movement direction corresponding with a visual direction in a stimulus than a different visual direction. Meanwhile, if there is no clear directional information in the visual flow, the response time will not be significantly different when a subject executes any gesture on a stimulus with no specific visual direction. Therefore, the hypotheses are established with reference to the relations between visual flows and touch gestures in the preliminary study (see Fig. 4) as follows:

Hypothesis 1. The mean response time for a slide gesture is significantly faster on stimuli that have a horizontally directed flow than stimuli that have a vertically directed flow.

Hypothesis 2. The mean response time for a scroll gesture is significantly faster on stimuli that have a horizontally directed flow than stimuli that have a vertically directed flow.

Hypothesis 3. The mean response times for both gestures are not significantly different between stimuli that have a horizontally directed flow than those with a vertically directed flow.

5.3. Method

5.3.1. Participants

Twenty-six undergraduate and graduate students (16 females, 10 males) in South Korea participated in this experiment. The participants ranged in age from 19 to 31 (M = 25.4, SD = 2.9). Except for one participant, all participants reported that they are right-hand dominant. The participants do not have any visuomotor impairment in the interaction with using digital devices. They were compensated with 3000 KRW for their participation.

5.3.2. Apparatus and stimulus

The stimuli were presented on a 27-inch monitor (with a display resolution set to 1280 × 720 pixels and a frame frequency of 144 Hz). The responses were collected when subjects released the right end key to a button box. The presentation of stimuli and collection of responses were controlled by PsychoPy3. The height of the monitor was approximately adjusted to the center of the fixation cross and the eye level of the subjects. The distance from the monitor to the subject's eyes was consistently about 60 cm.

In the experiment, there are two kinds of stimuli: screen and gesture. The set of screen stimuli was entirely the same as the set of stimuli in the preliminary study, thus consisting of a total of 16 stimuli (eight screen patterns with two modalities). The set of gesture stimuli consisted of two videos that include a semi-transparent photograph of a right hand straightening the index finger, which is sliding or scrolling. These screen-gesture stimuli were edited and merged into one video file in accordance with a below experimental procedure in Adobe After Effects 2020. The video files were finally rendered at a resolution of 1280 × 720 pixels at 30 frames per second (fps).

5.3.3. Tasks and procedure

Before the actual experiment, the subjects performed a practice session composed of 24 trials using the same experimental procedure; but the on-screen stimuli and gesture stimuli during practice consisted of random smart device screens such as a map or message applications irrelevant to the stimuli for the main experiment to avoid a practice effect.

Before the main experiment, the participants were asked to press the left end key. At the start of the experiment, the word 'next' was presented in Korean only once for 2 s. Then, the main experiment started with the first trial. Each trial started with a white fixation cross on a black background for 500 ms. After that, a blank screen appeared for 500 ms. Immediately, a 30 fps video for the screen and gesture stimuli played for 1500 ms. First, a screen stimulus, which is a static image, appeared in the center of the screen for eight frames (about 267 ms). Next, a moving semi-transparent hand gesture stimulus was overlapped and played on the screen stimulus for eight frames (about 267 ms). A beep sound was played at the moment the gesture stimulus disappeared. With the beep sound, the subjects were asked to say the word 'next' and to lift their index finger from the key and mimic the gesture toward the screen stimuli on the monitor as soon as possible. Response times were collected for 1000 ms from the beginning moment of the beep sound. After the screen and hand stimuli disappeared, a blank screen appeared for 1000 ms. Lastly, the subjects were asked to bring their hand back and to press the key again.

This trial was repeated 96 times for one experiment. In the entire trial, the screen and gesture stimuli were equally assigned in random order. To avoid passive responses neglecting stimuli, the message "focus on the screen stimuli" appeared twice, at the moment of one-third and two-thirds into the trial. The experimental procedure is schematized in Fig. 5.

5.4. Results

The data preprocessing proceeded in three steps. First, the data of two participants were excluded from the analysis. One participant responded incorrectly over 50% of the entire trial. Another participant was left-hand dominant, although we had recruited right-hand dominant participants. A left-hand dominance has been pointed out to be neurologically and behaviorally different to a right-hand dominance, thus they are traditionally excluded in numerous psychological experiments (Willems, Der Haegen, Fisher, & Francks, 2014). Second, incorrect responses were excluded from the entire dataset. There were two types of incorrect responses: the first one, which failed to accurately execute a proposed gesture, and the second one, which failed to lift their finger off the button box. It was only 0.02% of the entire data. Lastly, the mean response times (RTs) on each stimulus were calculated from the rest of the responses for each participant.

For overall understanding, a one-way analysis of variance (ANOVA) was first conducted for screen-gesture correspondence (corresponding/non-corresponding/neutral) on the total stimuli. The mean RTs among three correspondence conditions were not significantly different, $F(2,765) = 0.829$, $p = 0.437$; but the non-corresponding condition ($M = 0.343$, $SD = 0.104$) and neutral condition ($M = 0.344$, $SD = 0.105$) were about 10 ms longer than the corresponding condition ($M = 0.334$, $SD = 0.107$). These results indicate that only a part of the stimuli could show or none of them shows significant differences among corresponding conditions. Thus, we further conducted one-sided paired t-tests on the horizontal and vertical stimuli, and two one-sided paired t-tests on the neutral stimuli based on the three hypotheses. The results of the t-tests are summarized in Table 2.

Hypothesis 1 is partially supported at the 0.05 level of significance. Specifically, stimuli 1 b and 1c showed significant correspondence effects, whereas the other stimuli did not. More specifically, in stimulus 1a, the corresponding gesture slide (M = 0.335, SD = 0.116) was not significantly faster than the non-corresponding gesture scroll (M = 0.337, SD = 0.101). In stimulus 1 b, the corresponding gesture slide (M = 0.317, SD = 0.097) was 20 ms faster than the non-corresponding gesture scroll (M = 0.337, SD = 0.099); the size of the corresponding

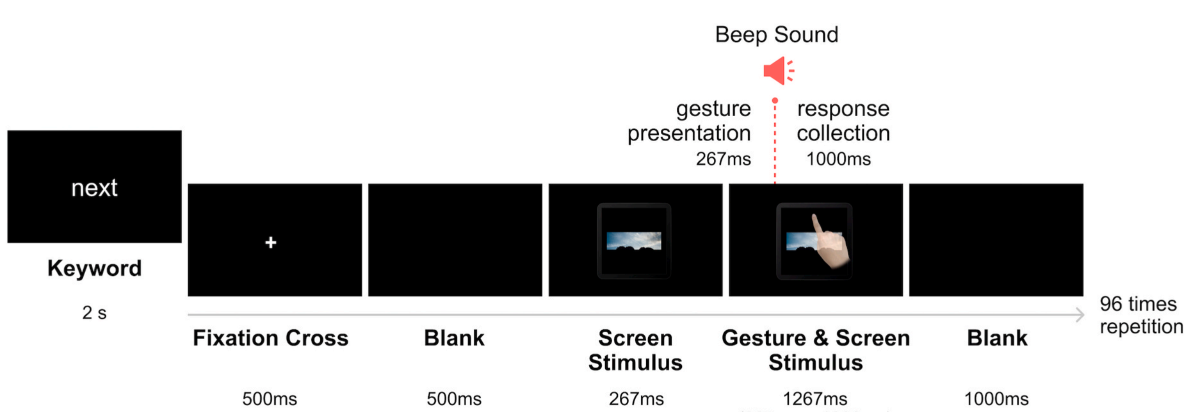

Fig. 5. Experimental procedure.

Table 2
Results of the paired samples t-tests. One-tailed t-tests were conducted for stimuli 1a to 2c, and two one-sided t-tests were conducted for stimuli 3a and 3b.

Stimulus	Gesture		t	df	*p*	Effect Size (Cohen's d)
	corresponding	non-corresponding				
1a	Slide	scroll	−0.180	47	0.429	−0.0260
1b	Slide	scroll	−2.066	47	0.022*	−0.2982
1c	Slide	scroll	−3.137	47	0.001***	−0.4527
2a	Scroll	slide	0.597	47	0.723	0.0861
2b	Scroll	slide	−0.959	47	0.171	−0.1384
2c	Scroll	slide	0.447	47	0.672	0.0646
3a	–	–	−0.184	47	0.855	−0.0266
3b	–	–	−0.638	47	0.527	−0.0921

$^{**}p < 0.01$; $^{*}p < 0.05$; $^{***}p < 0.001$.

effect was 0.02 (SD = 0.067). In stimulus 1c, the corresponding gesture slide (M = 0.325, SD = 0.106) was 33 ms faster than the non-corresponding gesture scroll (M = 0.359, SD = 0.115); the size of the corresponding effect was 0.333 (SD = 0.074).

Hypothesis 2 is rejected at the 0.05 level of significance. In stimulus 2a, the corresponding gesture slide (M = 0.357, SD = 0.105) was not significantly faster than the non-corresponding gesture scroll (M = 0.351, SD = 0.117). In stimulus 2 b, the corresponding gesture slide (M = 0.333, SD = 0.104) was not significantly faster than the non-corresponding gesture scroll (M = 0.343, SD = 0.090). In stimulus 2c, the corresponding gesture slide (M = 0.335, SD = 0.114) was not significantly faster than the non-corresponding gesture scroll (M = 0.332, SD = 0.100).

Hypothesis 3 is supported at the 0.05 level of significance. In stimulus 3a, the gesture slide (M = 0.348, SD = 0.101) was not significantly different from the gesture scroll (M = 0.349, SD = 0.108). To strictly confirm whether or not the two gestures are equivalent in response time, we further conducted a two-sample equivalence test with the equivalence bound of Cohen's $d_z = -0.5$ and $d_z = 0.5$. In the results, the difference between the two gestures is within its equivalence bound of −0.0334 and 0.0334 (95% confidence interval from −0.0212 to 0.0176). This result shows that the two gestures are significantly equivalent at the 0.05 level. In stimulus 3 b, the gesture slide (M = 0.338, SD = 0.105) was not significantly different from the gesture scroll (M = 0.344, SD = 0.108). We additionally conducted a two-sample equivalence test with the equivalence bounds of Cohen's $d_z = -0.5$ and $d_z = 0.5$. In the results, the difference between the two gestures is within its equivalence bound of −0.0379 and 0.0379 (95% confidence interval from −0.0290 to 0.0151). This result shows that the two gestures are significantly equivalent at the 0.05 level.

5.5. Discussion and implications

5.5.1. Discussion of findings

This main study investigates object-based correspondence effects between the visual flow of interface objects on the screen and a touch gesture. More specifically, we compare the participants' behavioral performance on the corresponding and non-corresponding gestures elicited by end-users in the preliminary study. Overall, the experimental results indicate that the effects of visual flow, the complex variable of shape, layout, number of entities, and other visual properties, is partially significant in gesture execution for the function 'next.' These results are mostly consistent with the results of a preliminary study on the neutral condition and the horizontal condition corresponding with the gesture 'slide', whereas they are inconsistent with the vertical condition corresponding with the gesture 'scroll.' To understand these findings more thoroughly, we discuss these results with individual visual properties for each stimulus in the comparison of the results of the preliminary study.

The overall results indicate that manipulating one entity and multiple entities are significantly different in the execution of a touch gesture at the perceptuomotor level. The number of entities, the individual variable, has consistently shown its significance in both the preliminary and main studies. In the preliminary study, the number of entities shows marginally significant differences in gesture selection between slide and scroll. Specifically, participants select slide (58.2%) slightly more than scroll (41.8%) on the stimuli with one entity, whereas slide (72.1%) is selected considerably more than scroll (27.9%) on the stimuli with multiple entities. This implies that users might independently perceive screen patterns with one entity and multiple entities to select an appropriate touch gesture. Although this result is significant only at the 0.1 level, the results are quite convincing because the effect of the number of entities is observed in the function 'view' and even in our previous study (Kim & Lee, 2020b). In the main study, none of the stimuli with one entity show a significant difference in the response time between slide and scroll, whereas some of the stimuli with multiple entities only in horizontal flow show significant differences in the response time between two gestures.

These results of both the preliminary and main studies demonstrate that only one on-screen object might be perceived to be manipulated in any direction regardless of its size or shape. This interpretation is quite convincing because a user's manipulation is not strictly constrained by a particular law in the digital environment as compared to the physical environment that is governed by gravity, friction, and other numerous physical laws (Terrenghi, Kirk, Sellen, & Izadi, 2007). A digital object might be perceived as matter with no mass in zero gravity, thus being freely manipulated on the screen in a user's perception. Meanwhile, multiple on-screen entities seem to be visually perceived to constrain their movements by themselves in accordance with their arrangements. This inference is reasonable that those entities might be perceptually grouped and suggest their virtual moving path in a particular direction under the Gestalt principles (Wagemans et al., 2012; Wertheimer, 1923). This means that the entities are visually perceived as obstacles to each other, thus not passing the space on where the entities are placed. Additionally, the multiple entities in our stimuli perfectly shared the same visual properties and aligned to each other, which means that the perceptually grouping effect is maximized.

This perceptuomotor difference in manipulating one entity and multiple entities is consistent with previous studies related to the limitations of direct manipulation. The direct manipulation of multiple objects has been pointed out to be a more challenging interaction for designers and users as compared to the manipulation for one object. Thus, several assistive technologies have been proposed, for example, for identifying multiple objects with different spatial properties as a group, selecting intended entities as a group, and manipulating the entities simultaneously (Frohlich, 1993; Kwon, Javed, Elmqvist, & Yi, 2011; Lindlbauer, Haller, Hancock, Scott, & Stuerzlinger, 2013; Xia, Araujo, & Wigdor, 2017). These studies convincingly demonstrate the significant effect of number of entities in various manners at the perceptuomotor level.

The effect of visual flow is not observed in the stimuli with the vertical condition corresponding with the gesture 'scroll' in the main study. This result seems to be inconsistent with the results of the preliminary study. However, this inconsistency is not unpredictable in terms of the priority of the primary gesture. In the preliminary study, despite equal numbers of the stimuli with each condition, the gesture 'slide' is selected overall twice as much as the gesture 'scroll.' This huge difference implies that the primary gesture 'slide' might be prioritized compared to another possible gesture 'scroll.' This priority might intervene with the effect of correspondence between vertical flow and scroll in the main study, thus deriving non-difference in response times between slide and scroll. To examine this hypothesis, we further analyzed an additional *t*-test to determine whether there was a difference in gesture execution between two gestures. The results show a significant difference in response time between the two gestures (two-tailed paired *t*-test, $t = -2.24$, $p = 0.026$). The mean response time on the gesture 'slide' is significantly faster than the gesture 'scroll.' This result also demonstrates the priority of the primary gesture representation as well as the preliminary study. This priority can be interpreted, abovementioned, to be derived from the direction inherent in the text and image content. If the experiment had suggested stimuli consisting of entities rotated 90°, the results would have been entirely different.

According to all the above discussions, there is the effect of on-screen affordances for touch gestures, but it shows different manners from physical affordances. This difference might be derived from different interaction modalities grounded on each physical and digital environment. Specifically, the manipulation in a physical environment is freely executed in 3-dimensional space depending complexly on visual, auditory, and tactile modalities, whereas the manipulation in a digital environment, particularly on a touchscreen device, is limitedly allowed in the 2-dimensional space on the screen depending mostly on the visual modality. This means that an action possibility—affordance is determined by environment-innate manipulable structural characteristics, which are perceptible by specific modalities. Thus, although the touchscreen cannot provide tactile impressions of an on-screen object except the quality of the glass panel in the screen, depicted pixels can potentiate enough real touch gestures grounded on the capabilities of devices, hardware, and software. In this context, we can conclude that designing a touch gesture should be investigated under the digital materiality of the interface.

5.5.2. Implications

The present study provides theoretical and methodological implications for understanding and evaluating the digital affordance of an interface object with the touch gesture. First, the results of our study substantiated the ontological allegation of the on-screen affordance (Norman, 2013). The results demonstrated that the digital affordance has a significant effect of facilitating the touch gesture based on the interface object's perceptible properties for a particular function at the perceptuomotor level. This finding suggests that the interface design on the touchscreen should consider the intuitive visual design but also the appropriate touch gesture design for supporting the user's affordance perception. Under this theoretical view, this study additionally provides a new methodological approach for gesture design. The gestural studies have focused on merely geometric and kinematic characteristics (e.g., body part, pose, flow) of the user-elicited gesture itself (Vogiatzidakis & Koutsabasis, 2018; Vuletic et al., 2019, 2021), neglecting the effect of the digital affordance on the screen. However, our study proposed an empirical method to evaluate the intuitiveness of user-elicited gestures through examining an object-based correspondence—affordance effect. This method enables researchers to objectively investigate the genuine effect of digital affordance beyond subjective evaluations of user-elicited gestures (Dingler, Rzayev, Shirazi, & Henze, 2018; Villarreal-Narvaez, Vanderdonckt, Vatavu, & Wobbrock, 2020).

In addition, this study provides managerial and practical implications for digital services on the touchscreen. The results of this study demonstrate that an interface design considering touch gestures can support a user's purposeful and seamless experience in the mobile device. Specifically, our study investigated the affordance effect of various design patterns which present one or multiple contents on the screen. These design patterns can be easily found in most e-commerce mobile websites or applications for product assortment. These online assortment patterns have been considered to affect customers' perception, moreover, their purchase intention (Kahn, 2017). On these design patterns, customers perform touch gestures with intentions to navigate, purchase products and other purposeful behavior (Shi & Kalyanam, 2018). In this context, this study emphasizes the requirement to support customers to accomplish their goals through various design patterns of online assortment corresponding to touch gestures. Moreover, the results of our study provide practical design guidelines in detail to develop these patterns for online retailers. The well-designed design patterns and touch gestures will increase not only customers' utilitarian experience but also hedonic experience, thereby increasing user engagement (Dou & Sundar, 2016; Sundar, Bellur, Oh, Xu, & Jia, 2014; Wang & Sundar, 2018).

5.5.3. Limitations

There are some limitations to this study. First, the main study only explores the function 'next.' Although this function was selected in accordance with the results of the preliminary study, another function 'view' and other possible functions might produce different results based on their embodied schema. According to the conceptual metaphor theory (Lakoff & Johnson, 2003), a concept is specifically established by relevant embodied experiences and further extensively associated with the abstract concepts. This demonstrates that the function 'next' is constituted of sensorimotor experiences on surrounding sides in reference to an observer's body. In this context, a pragmatic difference between some functions might derive different perceptual effects, and we could have obtained more abundant insights on digital affordance. Thus, a future study will investigate the affordance effect on various functions. Next, the stimuli are limited to the text and image modalities. These two modalities were selected because of their commonness and the convenience to manipulate their visual properties as compared to other interface objects. These text and image-based objects have the same characteristic aiming to deliver content for users. Yet, other interface objects have different functions to navigate or control the interface, for example, a button, scroll bar, or slider. Thus, other interface objects will be investigated in our future study.

6. Conclusion

This study investigates the effect of digital affordances on the screen to examine an object-based correspondence effect for user-defined touch gestures on the screen design pattern manipulated in several visual properties. In the results, the effect of on-screen affordance is observed in the participant's touch gesture execution, but occurs in a different manner than physical affordance. This difference is attributed to the difference in interactional modalities between physical and digital environments. More specifically, the on-screen affordance is established on manipulable structures emerging out of the technological specifications of both hardware and software. This emergence consequently produces unique characteristics grounded on the digital environment. The characteristics of on-screen affordances are summarized as follows:

1) A digital entity is perceived as freely manipulable regardless of physical laws.
2) A visuo-perceptible entity can only be perceptually constrained by other visual obstacles.
3) The gestural priority depends on the embodied direction inherent in particular content.

These are only a part of the characteristics of on-screen affordances derived from the results on the function 'next.' We believe that the digital affordances on the touchscreen are still in uncharted territory, having left many questions on their characteristics. However, the present study successfully found out that the on-screen affordance significantly influences users' touch gesture execution. This conclusion suggests that the user-defined gesture should be investigated in accordance with an effect of digital affordances based on the hybrid materiality consisting of both characteristics of software and hardware of interactive media. We expect that these kinds of characteristics on touchable pixels will be applicable for designing a more intuitive on-screen object and appropriate gestures.

Credit author statement

Sangyeon Kim took charge of the conceptualization, stimuli design, data collection, Formal analysis, writing, and manuscript revisions. Sangwon Lee helped the conceptualization, Formal analysis, writing and revising a manuscript, and contributed to funding acquisition.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2022-0-00078, Explainable Logical Reasoning for Medical Knowledge Generation).

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1A2C4002026).

References

Allen-Robertson, J. (2018). Critically assessing digital documents: Materiality and the interpretative role of software. *Information, Communication & Society, 21*(11), 1732–1746. https://doi.org/10.1080/1369118X.2017.1351575

Azaad, S., & Laham, S. M. (2019). Sidestepping spatial confounds in object-based correspondence effects: The Bimanual Affordance Task (BMAT). *Quarterly Journal of Experimental Psychology, 72*(11), 2605–2613. https://doi.org/10.1177/1747021819852216

Berriman, L., & Mascheroni, G. (2019). Exploring the affordances of smart toys and connected play in practice. *New Media & Society, 21*(4), 797–814. https://doi.org/10.1177/1461444818807119

Bub, D. N., & Masson, M. E. J. (2010). Grasping beer mugs: On the dynamics of alignment effects induced by handled objects. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2), 341–358. https://doi.org/10.1037/a0017606

Bub, D. N., Masson, M. E. J., & Kumar, R. (2018). Time course of motor affordances evoked by pictured objects and words. *Journal of Experimental Psychology: Human Perception and Performance, 44*(March), 53–68. https://doi.org/10.1037/xhp0000431

Bub, D. N., Masson, M. E., & van Noordenne, M. (2021). Motor representations evoked by objects under varying action intentions. *Journal of Experimental Psychology: Human Perception and Performance, 47*(1), 53. https://doi.org/10.1037/xhp0000876

Chen, J., & Proctor, R. W. (2012). Up or down: Directional stimulus-response compatibility and natural scrolling. In *Proceedings of the human factors and ergonomics society 56th annual meeting* (pp. 1381–1385). https://doi.org/10.1177/1071181312561394

Chen, J., & Proctor, R. W. (2013). Response–effect compatibility defines the natural scrolling direction. *Human Factors, 55*(6), 1112–1129. https://doi.org/10.1177/0018720813482329

Chong, I., & Proctor, R. W. (2020). On the evolution of a radical concept: Affordances according to Gibson and their subsequent use and development. *Perspectives on Psychological Science, 15*(1), 117–132. https://doi.org/10.1177/1745691619868207

Cho, D. T., & Proctor, R. W. (2010). The object-based simon effect : Grasping affordance or relative location of the graspable part. *Journal of Experimental Psychology, 36*(4), 853–861. https://doi.org/10.1037/a0019328

Dingler, T., Rzayev, R., Shirazi, A. S., & Henze, N. (2018). Designing consistent gestures across device types: Eliciting RSVP controls for phone, watch, and glasses. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–12). https://doi.org/10.1145/3173574.3173993

Dou, X., & Sundar, S. S. (2016). Power of the swipe: Why mobile websites should add horizontal swiping to tapping, clicking, and scrolling interaction techniques. *International Journal of Human-Computer Interaction, 32*(4), 352–362. https://doi.org/10.1080/10447318.2016.1147902

Frohlich, D. M. (1993). The history and future of direct manipulation. *Behaviour & Information Technology, 12*(6), 315–329. https://doi.org/10.1080/01449299308924396

Galvao, A. B., & Sato, K. (2005). Affordances in product architecture: Linking technical functions and users' tasks. In *Volume 5a: 17th International conference on design theory and methodology* (pp. 143–153). https://doi.org/10.1115/DETC2005-84525, 5a.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gibson, E. J. (2000). Perceptual learning in development: Some basic concepts. *Ecological Psychology, 12*(4), 295–302. https://doi.org/10.1207/S15326969ECO1204_04

Hoffmann, E. R., & Chan, A. H. (2018). Review of compatibility and selection of multiple lever controls used in heavy machinery. *International Journal of Industrial Ergonomics, 65*, 93–102. https://doi.org/10.1016/j.ergon.2017.07.007

Hoffmann, E. R., Chan, A. H., Man, S. S., & Chan, L. C. (2019). Determining the validity of the visual field principle for designing control/display arrangements. *Applied Ergonomics, 81*, Article 102887. https://doi.org/10.1016/j.apergo.2019.102887

Hornecker, E. (2012). Beyond affordance: Tangibles' hybrid nature. In *Proceedings of the sixth international conference on tangible, embedded and embodied interaction* (pp. 175–182). https://doi.org/10.1145/2148131.2148168

Hurtienne, J., & Blessing, L. (2007). Design for intuitive use - testing image schema theory for user interface design. Proceedings of ICED 2007. In *The 16th international conference on engineering design* (pp. 1–12).

Janczyk, M., Xiong, A., & Proctor, R. W. (2019). Stimulus-response and response-effect compatibility with touchless gestures and moving action effects. *Human Factors, 61*(8), 1297–1314. https://doi.org/10.1177/0018720819831814

Jung, H., Wiltse, H., Wiberg, M., & Stolterman, E. (2017). Metaphors, materialities, and affordances: Hybrid morphologies in the design of interactive artifacts. *Design Studies, 53*, 24–46. https://doi.org/10.1016/j.destud.2017.06.004

Kahn, B. E. (2017). Using visual design to improve customer perceptions of online assortments. *Journal of Retailing, 93*(1), 29–42. https://doi.org/10.1016/j.jretai.2016.11.004

Kim, S., & Lee, S. (2018). Three levels of perceiving affordance. In *Proceedings of the 30th Australian conference on computer-human interaction* (pp. 261–265). https://doi.org/10.1145/3292147.3292219

Kim, S., & Lee, S. (2020a). Smash the dichotomy of Skeuomorphism and flat design: Designing an affordable interface to correspond with the human perceptuomotor process. *International Journal of Human-Computer Studies, 141*, Article 102435. https://doi.org/10.1016/j.ijhcs.2020.102435

Kim, S., & Lee, S. (2020b). Touch digitality: Affordance effects of visual properties on gesture selection. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–8). https://doi.org/10.1145/3334480.3382914

Kim, K. J., Sundar, S. S., & Park, E. (2011). The effects of screen-size and communication modality on psychology of mobile device users. In *Proceedings of the 2011 annual conference extended abstracts on human factors in computing systems* (pp. 1207–1212). https://doi.org/10.1145/1979742.1979749

Kwon, B. C., Javed, W., Elmqvist, N., & Yi, J. S. (2011). Direct manipulation through surrogate objects. In *Proceedings of the 2011 annual conference on human factors in computing systems* (pp. 627–636). https://doi.org/10.1145/1978942.1979033

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226470993.001.0001

Lindlbauer, D., Haller, M., Hancock, M., Scott, S. D., & Stuerzlinger, W. (2013). Perceptual grouping: Selection assistance for digital sketching. In *Proceedings of the 2013 ACM international conference on interactive tabletops and surfaces* (pp. 51–60). https://doi.org/10.1145/2512349.2512801

Maier, J. R. A., & Fadel, G. M. (2003). Affordance-based methods for design. In *ASME international design engineering technical conferences and computers and information in engineering conference* (pp. 785–794). https://doi.org/10.1115/DETC2003/DTM-48673

Masoudi, N., Fadel, G. M., Pagano, C. C., & Elena, M. V. (2019). A review of affordances and affordance-based design to address usability. In *Proceedings of the design society: international conference on engineering design* (pp. 1353–1362). https://doi.org/10.1017/dsi.2019.141, 1(1).

McGrenere, J., & Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of the graphics interface 2000* (pp. 179–186). https://doi.org/10.20380/GI2000.24

Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., Schraefel, C., et al. (2014). Reducing legacy bias in gesture elicitation studies. *Interactions, 21*(3), 40–45. https://doi.org/10.1145/2591689

Morris, M. R., Wobbrock, J. O., & Wilson, A. D. (2010). Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010* (pp. 261–268).

Nielsen, J. (2009). *Powers of 10: Time scales in user experience*. Nielsen Norman Group. https://www.nngroup.com/articles/powers-of-10-time-scales-in-ux/.

Norman, D. A. (1988). *The psychology of everyday things*. Basic books.

Norman, D. A. (2008). Signifiers, not affordances. *Interactions, 15*(6), 18–19. https://doi.org/10.1145/1409040.1409044

Norman, D. A. (2013). *The design of everyday things: Revised and* (expanded edition). Basic Books.

Pavese, A., & Buxbaum, L. J. (2002). Action matters: The role of action plans and object affordances in selection for action. *Visual Cognition, 9*(4–5), 559–590. https://doi.org/10.1080/13506280143000584

Phillips, J. G., Triggs, T. J., & Meehan, J. W. (2005). Forward/up directional incompatibilities during cursor placement within graphical user interfaces. *Ergonomics, 48*(6), 722–735. https://doi.org/10.1080/00140130500095795

Proctor, R. W., Lien, M. C., & Thompson, L. (2017). Do silhouettes and photographs produce fundamentally different object-based correspondence effects? *Cognition, 169*, 91–101. https://doi.org/10.1016/j.cognition.2017.08.009

Proctor, R. W., & Miles, J. D. (2014). Does the concept of affordance add anything to explanations of stimulus-response compatibility effects?. In*Psychology of learning and motivation - advances in research and theory* (1st ed., Vol. 60). Elsevier Inc. https://doi.org/10.1016/B978-0-12-800090-8.00006-8.

Proctor, R. W., & Vu, K. P. L. (2016). Principles for designing interfaces compatible with human information processing. *International Journal of Human-Computer Interaction, 32*(1), 2–22. https://doi.org/10.1080/10447318.2016.1105009

Proverbio, A. M., Adorni, R., & D'Aniello, G. E. (2011). 250 Ms to code for action affordance during observation of manipulable objects. *Neuropsychologia, 49*(9), 2711–2717. https://doi.org/10.1016/j.neuropsychologia.2011.05.019

Sharma, A., Roo, J. S., & Steimle, J. (2019). Grasping microgestures: Eliciting single-hand microgestures for handheld objects. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13). https://doi.org/10.1145/3290605.3300632

Shi, S. W., & Kalyanam, K. (2018). Touchable apps: Exploring the usage of touch features and their impact on engagement. *Journal of Interactive Marketing, 44*, 43–59. https://doi.org/10.1016/j.intmar.2018.06.001

Soni, N., Gleaves, S., Neff, H., Morrison-Smith, S., Esmaeili, S., Mayne, I., et al. (2020). Adults' and children's mental models for gestural interactions with interactive spherical displays. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–12). https://doi.org/10.1145/3313831.3376468

Steiner, L., Burgess-Limerick, R., & Porter, W. (2014). Directional control-response compatibility relationships assessed by physical simulation of an underground bolting machine. *Human Factors, 56*(2), 384–391. https://doi.org/10.1177/0018720813490953

Still, J. D., Still, M. L., & Grgic, J. (2015). Designing intuitive interactions: Exploring performance and reflection measures. *Interacting with Computers, 27*(3), 271–286. https://doi.org/10.1093/iwc/iwu046

Sundar, S. S., Bellur, S., Oh, J., Xu, Q., & Jia, H. (2014). User experience of on-screen interaction techniques: An experimental investigation of clicking, sliding, zooming, hovering, dragging, and flipping. *Human-Computer Interaction, 29*(2), 109–152. https://doi.org/10.1080/07370024.2013.789347

Sundar, S. S., Xu, Q., & Bellur, S. (2010). Designing interactivity in media interfaces. In *Proceedings of the 28th international conference on human factors in computing systems* (pp. 2247–2256). https://doi.org/10.1145/1753326.1753666

Terrenghi, L., Kirk, D., Sellen, A., & Izadi, S. (2007). Affordances for manipulation of physical versus digital media on interactive surfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1157–1166). https://doi.org/10.1145/1240624.1240799

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 830–846. https://doi.org/10.1037/0096-1523.24.3.830

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*(6), 769–800.

Vainio, L., Symes, E., Ellis, R., Tucker, M., & Ottoboni, G. (2008). On the relations between action planning, object identification, and motor representations of observed actions and objects. *Cognition, 108*(2), 444–465. https://doi.org/10.1016/j.cognition.2008.03.007

Vatavu, R.-D., & Wobbrock, J. O. (2015). Formalizing agreement analysis for elicitation studies. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 1325–1334). https://doi.org/10.1145/2702123.2702223

Vermeulen, J., Luyten, K., van den Hoven, E., & Coninx, K. (2013). Crossing the bridge over norman's gulf of execution. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1931–1940).

Villarreal-Narvaez, S., Vanderdonckt, J., Vatavu, R. D., & Wobbrock, J. O. (2020). A systematic review of gesture elicitation studies: What can we learn from 216 studies?. In *Proceedings of the 2020 ACM designing interactive systems conference* (pp. 855–872). https://doi.org/10.1145/3357236.3395511

Vogiatzidakis, P., & Koutsabasis, P. (2018). Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies and Interaction, 2*(4), 65. https://doi.org/10.3390/mti2040065

Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., & Grealy, M. (2019). Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies, 129*, 74–94. https://doi.org/10.1016/j.ijhcs.2019.03.011

Vuletic, T., Duffy, A., McTeague, C., Hay, L., Brisco, R., Campbell, G., et al. (2021). A novel user-based gesture vocabulary for conceptual design. *International Journal of Human-Computer Studies, 150*, Article 102609. https://doi.org/10.1016/j.ijhcs.2021.102609

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., et al. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin, 138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wang, R., & Sundar, S. S. (2018). How does parallax scrolling influence user experience? A test of time (theory of interactive media effects). *International Journal of Human-Computer Interaction, 34*(6), 533–543. https://doi.org/10.1080/10447318.2017.1373457

Wertheimer, M. (1923). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-005.

Willems, R. M., Der Haegen, L. Van, Fisher, S. E., & Francks, C. (2014). On the other hand: Including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience, 15*(3), 193–201. https://doi.org/10.1038/nrn3679

Williams, A. S., & Ortega, F. R. (2020). Evolutionary gestures. *Interactions, 27*(5), 50–53. https://doi.org/10.1145/3412499

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the 27th international conference on human factors in computing systems - CHI 09* (pp. 1083–1092). https://doi.org/10.1145/1518701.1518866

Worringham, C. J., & Beringer, D. B. (1998). Directional stimulus-response compatibility: A test of three alternative principles. *Ergonomics, 41*(6), 864–880. https://doi.org/10.1080/001401398186694

Xia, H., Araujo, B., & Wigdor, D. (2017). Collection objects: Enabling fluid formation and manipulation of aggregate selections. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 5592–5604). https://doi.org/10.1145/3025453.3025554