Pen + mid-air gestures: eliciting contextual gestures

Ilhan Aslan, Tabea Schmidt, Jens Woehrle, Lukas Vogel, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Aslan, Ilhan, Tabea Schmidt, Jens Woehrle, Lukas Vogel, and Elisabeth André. 2018. "Pen + mid-air gestures: eliciting contextual gestures." In *Proceedings of the 2018 on International Conference on Multimodal Interaction - ICMI '18, Boulder, CO, USA, October 16 - 20, 2018*, edited by Sidney K. D'Mello, Panayiotis (Panos) Georgiou, Stefan Scherer, Emily Mower Provost, Mohammad Soleymani, and Marcelo Worsley, 135–44. New York, NY: ACM Press. https://doi.org/10.1145/3242969.3242979.

Pen + Mid-Air Gestures: Eliciting Contextual Gestures

Ilhan Aslan, Tabea Schmidt, Jens Woehrle, Lukas Vogel, and Elisabeth André
Human-Centered Multimedia Lab, Augsburg University
Augsburg, Germany
lastname@hcm-lab.de

ABSTRACT

Combining mid-air gestures with pen input for bi-manual input on tablets has been reported as an alternative and attractive input technique in drawing applications. Previous work has also argued that mid-air gestural input can cause discomfort and arm fatigue over time, which can be addressed in a desktop setting by allowing users to gesture in alternative restful arm positions (e.g., elbow rests on desk). However, it is unclear if and how gesture preferences and gesture designs would be different for alternative arm positions. In order to inquire these research question we report on a user and choice based gesture elicitation study in which 10 participants designed gestures for different arm positions. We provide an in-depth qualitative analysis and detailed categorization of gestures, discussing commonalities and differences in the gesture sets based on a "think aloud" protocol, video recordings, and self-reports on user preferences.

CCS CONCEPTS

• **Human-centered computing → Interaction techniques**;

KEYWORDS

Bi-Manual input, touch, pen, gesture elicitation

ACM Reference Format:
Ilhan Aslan, Tabea Schmidt, Jens Woehrle, Lukas Vogel, and Elisabeth André. 2018. Pen + Mid-Air Gestures: Eliciting Contextual Gestures. In *2018 International Conference on Multimodal Interaction (ICMI '18), October 16–20, 2018, Boulder, CO, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3242969.3242979

1 INTRODUCTION

Today's state of the art touch screen devices, such as Apple's iPad Pro or Microsofts' Surface models support dedicated and precise pen-based input in addition to multi-touch (and thus bi-manual input). The input of digital pens can easily be combined and complemented with other modalities during bi-manual input, allowing users, for example to use one hand to draw a line with a pen and change the ink color with their other hand.

Aslan et al. [3] have recently demonstrated the attractiveness of pen and mid-air input on tablets as a complementary input modality to pen and multi-touch gestures. Furthermore, they designed a set of mid-air gestures for operations in drawing applications, such as zoom canvas, rotated canvas, and open drawing tools, which they elicited from users in a participatory design process.

Despite the potential of pen and mid-air input to expand the interaction landscape on interactive surfaces to the better, and its potential to address interaction issues, such as occlusion on small sized touch screens, the use of mid-air gestures is in many situations critical. The colloquial term "the gorilla arm syndrome" is often used to refer to issues with fatigue and discomfort when postures have to be maintained over long periods of time and gestures have to be performed in a repetitive manner (e.g., [18, 37]).

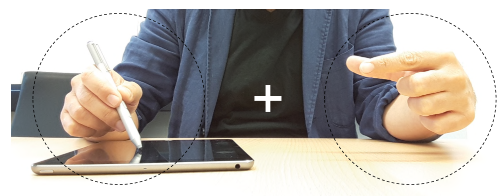

Figure 1: Illustration of pen and mid-air input on a tablet. Note that the elbow of the non-preferred hand (NP) rests on the desk during the mid-air gestural interaction.

Related research (e.g., [16, 19]) has argued that allowing users to rest their arm (e.g., as illustrated in Figure 1) while performing hand gestures would limit the inherent effects of gestural mid-air interfaces. We believe that enabling gestural interaction in different arm positions would provide users with flexibility and physical comfort (i.e., they wouldn't have to maintain the same posture over long periods of time), and thus, allow them to work longer periods of time before fatigue actually hits and a break may be required.

However, it is unclear how users would perform gestures in restful arm positions compared to the "non-rest" position; and if they would prefer the same gestures for triggering the same commands in different arm positions? In order to address this research question, we build on previous research [3] and report on a user and choice-based elicitation study with ten participants. For the same 13 commands (e.g., zoom, change ink size) gestures were elicited in two rest positions (i.e., elbow rests on desk, and arm rests on desk) and a non-rest position. The sets of gestures for the three different positions are analyzed in detail and guidelines are presented for the design of contextual/restful gestures for combined pen/touch and mid-air input on tablets.

2 RELATED RESEARCH

The benefits of using multi-touch and pen input in tandem on large interactive surfaces, such as tabletops are well explored (e.g.,

ICMI '18, October 16–20, 2018, Boulder, CO, USA © 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5692-3
https://doi.org/10.1145/3242969.3242979

[11, 17, 23, 26, 43, 45]). However, with small-sized interactive surfaces, such as with tablets bi-manual touch input faces more constraints. For example, on top of screen sizes being insufficient, it is easier to cause undesired screen occlusion. Yoon et al. [45] have studied how pen and touch can be used in combination to compensate insufficient space on a tablet device for making annotations by tearing the digital paper and creating additional space. An approach to overcome interaction constraints with uni-modal solutions without introducing additional user interactions is the integration of alternative input modalities.

Previous research has already studied, for example, tactile and tangible [10], voice and tactile [6], tactile and pressure [13], or pressure and tangible input [9]. Today's sensing technologies are enabling low-cost hand motion systems that are capable to transform "any" mid-air space into sensory space for gestural interaction. A resulting trend is expanding the interaction space of touch screen devices, such as mobiles and tablets, above and around the actual touch-sensitive screens (e.g., [2, 22, 27, 44, 46]). This trend shows promise to address many of the limitations in unimodal touch interaction with interactive surfaces by supporting alternative input techniques and combining mid-air input with touch input. However, the expanded interaction space poses a challenge for gestural interaction design. A mismatch between designed gestures and how real users would want to use gestures in interactions (i.e., the gulf of execution [29]) is one of the issues designers and developers encounter when designing gestural interaction for novel and future interaction spaces. In order to reduce a potentially large gulf of execution many researchers have adopted the user-elicitation method by Wobbrock et al. [42], which is based on the guessability technique [28, 40].

2.1 Gesture Elicitation Studies

The aim of gesture elicitation studies is to design gestures for a predefined set of operations, which are represented during the studies by referents. Referents are usually (visual) effects (i.e., video recordings or animations), demonstrating on the target system the transition between a pre-gesture and post-gesture state of the system. The effects are often accompanied with written descriptions of the operations. For example, for a zoom operation on a touch screen, a referent would show the effect of the zoom operation on the touch screen without indicating how the effect is accomplished by the user. Participants of elicitation studies are first asked to watch the visual referents (and read the textual descriptions) and then to demonstrate ("guess") a gesture to accomplish the presented effect. Participants' demonstrations and their explanations of why they would use a specific gesture to accomplish a specific effect are video recorded. The recordings are then used to establish gesture taxonomies based on participants' demonstrations and to define/design a resulting gesture set. Wobbrock et al. [40] have proposed a method to evaluate guessability by computing agreement scores over participants, referents, and proposed gestures. These agreement scores are measures, explaining how well (future) users will guess specific gestures to accomplish specific effects. In a later work, Wobbrock et al. [42] have applied agreement scores to elicit a set of surface gestures.

The effectiveness of user elicitation studies has motivated many researcher to adopt the approach in order to elicit gestures for various target systems and novel interaction settings, such as TVs (e.g., [36]), phones (e.g., [32]), robots (e.g., [30]), virtual characters (e.g., [24]), wall-sized displays (e.g., [39]), augmented reality applications (e.g., [31]), elastic deformable displays (e.g.,[34]), smart glasses (e.g., [35]), and for pen and mid-air input on tablets [3].

In summary, gesture elicitation studies have been applied mainly to design gesture sets for novel interfaces and new design spaces. Only a few of recent works have adapted the user elicitation method with choice-based elicitation (e.g., [14, 33]), either to build expand existing gesture sets or to provide guidance for specific user groups (e.g., blind users) and application scenarios (e.g., intense gameplay).

In contrast to previous work in elicitation studies, this paper aims to explore not a novel interface, but relevant (new) interaction contexts. To this end, the paper utilizes existing gesture choices for the default (technology-centered) context to explore how users would want to gesture in rest positions, which either result from fatigue and discomfort or (more importantly) are taken as a precaution of users to prevent fatigue. Next we provide arguments based on related work for why fatigue and rest positions are of relevance for the design of mid-air gestures.

2.2 Mid-Air Input and Fatigue

Earlier research in tabletop interaction design (e.g., [20, 38]) has already recognized the constraints of planar interaction spaces, and thus, the space above the tabletop and how it could be used to add depth and continuity to interactions has been studied. For example, Marquardt et al. [25] explored combining touch on tabletops and mid-air gestures above tabletop screens towards a continuous interaction space for tabletop interaction. Annett et al. [1] made use of Proximity information to contextualize tabletop interaction, such as distinguishing and adapting to left and right hand usage.

More recently, mid-air and touch input has been proposed for various other devices and contexts, in which combined input is proposed to address contextual constraints, such as target acquisition performance on touch screens in cars [4] and collaboration settings [5]. Others have suggested to address typical issues with small-sized screens and occlusion. Chen et al. [12], for example, suggested to combine pre-touch or post-touch gestures with touch allowing alternative forms to perform operations, such as to zoom by tabbing the screen followed by circling with the finger above the screen. Hinckley et al. [22] suggested to adapt touch interfaces depending on the posture of the approaching hand, for example to differentiate a two-finger-zoom intention from a thumb-tab intention on the screen and thus reduce information overload on the screen. Motivated by recent advances in pen-based input, mid-air input has also been proposed to complement pen input on tablets with mid-air gestures for (bi-manual) pen and mid-air input [3].

While there are potential benefits of mid-air interfaces for different use contexts [37], there are also common and systematic constraints of off-the-shelf controllers, such as the Xbox Kinect or the Leap Motion. As optical systems they have specific requirements on orientation and proximity towards the inbuilt sensors. Thus, they require users to perform gestures in a limited and invisible sensory space, causing discomfort and fatigue, because postures have to be maintained over long periods of time and performed in a repetitive manner (e.g., [18, 37]). Furthermore, fatigue is influenced by physiological and psychological factors [15], and thus professional graphic artists and designers, who are known to work ad irregular times and for long hours are specifically prone to arm fatigue. With gestural interaction becoming more popular, HCI research has re-addressed arm fatigue as a dilemma in interaction design. Thus, researchers are presenting insights towards a better understanding and ways to quantify, model, and predict which gestures may reduce arm fatigue (e.g., [7, 8, 18, 21]). Others (e.g., [16, 19]) have suggested that allowing users to rest their arms while performing hand gestures would already limit the inherent effects of gestural mid-air interfaces.

We believe the relevance for fatigue and fatigue prevention in interaction design increases with more technologies enabling physical and body-based interactions in new situations. Preventing fatigue may influence user performance over time and have a positive impact on user acceptance and experience. We contribute to ongoing efforts in HCI to prevent fatigue by adopting and adapting a user and choice based elicitation study to design a contextual gesture set (i.e., a set of gestures, which can be performed in alternative postures and restful arm positions).

3 ELICITATION STUDY

We have conducted a gesture elicitation study to achieve our overall aim of addressing the interaction space for pen and mid-air gestural input on tablets as a continuous space where users may (want to) switch between different postures/arm positions; i.e., (i) a non-rest position, (ii) elbow rests on the desk, and (iii) arm rests on the desk.

3.1 Participants and Apparatus

Ten participants (8m, 2f) took part in the gesture elicitation study. All participants were recruited at the university and thus were either university staff or students. Half of the participants were left handed. Two researchers conducted the study in an empty office with one researcher providing instructions and the other setting up the camera and taking notes. Each session took about one hour. We choose to use the same number of participants as Aslan et al [3], but compared to the previous work we applied a repeated observations method, which resulted in longer sessions (about twice as long) with each participant.

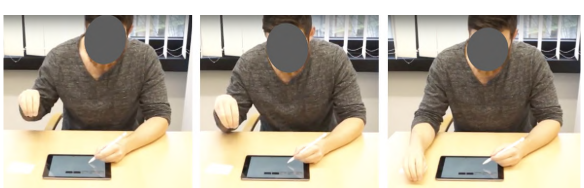

Figure 2: Exemplary images of a participant trying out a gesture in different arm positions for his non-preferred hand (NP) while using a pen in his preferred hand (P).

During the introduction of the study participants could try out an exemplary drawing application (i.e., *procreate*[1]) on an iPad Pro using Apple's iPencil. They could try out the application for as long as they needed to get a good grasp for the application and task context. Afterwards, participants were informed that related research had already proposed mid-air gestures as an alternative input modality for (bi-manual) multi-touch input; and that the aim of the study is to design gestures for alternative arm positions, which users may take for physical comfort and to prevent arm fatigue. Figure 2 presents the setup and example screenshots of a participant exploring how performing a gesture feels in the three different arm positions.

3.1.1 Referents.

Two separate sets of referents were prepared for the gesture elicitation study. Table 1 explains the referents. The first set consisted of referents for eight operations, for which related work had already elicited mid-air gestures for NP in the non-rest position. Thus, for these operations we have "implemented" the existing gesture choices as video recordings for the non-rest positions. Figure 3 shows screenshots from the gesture choice videos, which are annotated with sketches to illustrate properly each gesture as an image. For the preferred hand we chose to implement only finger touch and pen touch versions (and discard "pen hover" as a possible alternative for finger touch), since state of the art devices can already differentiate between finger and pen touch.

ReferentID - Operation/Referent name	*SetID - Gesture choices*
1 - Pan: Move canvas	1 - 2a,2b
2 - Change ink size of pen	1 - 1a,1b,1c
3 - Rotate canvas	1 - 3a, 3b
4 - Change ink color of pen	1 - 5a
5 - Open/Close menu	1 - 5b
6 - Change brush	1 - 5a
7 - Undo/Redo	1 - 6a,6b,6c
8 - Zoom	1 - 4a,4b
9 - Draw&change ink size	2 - no choice
10 - Draw&change ink transparency	2 - no choice
11 - Switch to color picker (pipette)	2 - no choice
12 - Draw&change ink color	2 - no choice
13 - Switch to eraser	2 - no choice

Table 1: The list of 13 operations for which participants received referents and proposed gestures. Each referent was represented as an animation/video and a textual description. Two sets of referents were used. For the first set of gestures choices were available as video recordings (see Figure 3). For the second set no choices were available.

The second set consisted of referents for five operations for which no gesture choices existed based on previous work. Three of the operations are variations of an operation in the first set which have to be performed while drawing. The last two operations (i.e., switch to eraser, and pick a color with a color-picker) are new and we added those since in our own experience (and based on an exploration of exiting drawing applications) they are standard operations in drawing applications.

[1] http://procreate.si/

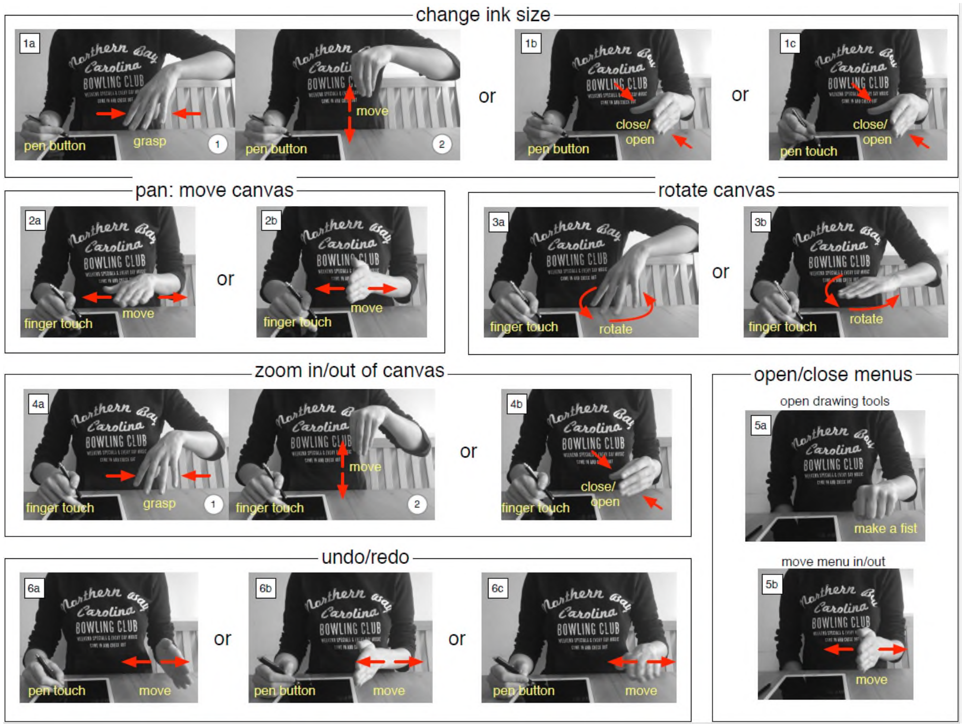

Figure 3: Overview of the gestures choices that were presented as gesture videos to the participants.

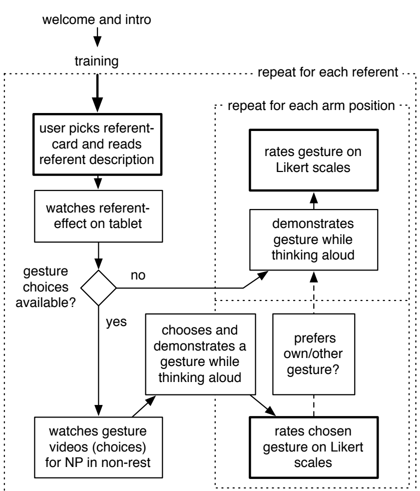

Figure 4: Procedure for the user and choice based elicitation study

3.2 Procedure

The gesture elicitation procedure started with participants choosing a referent. Figure 4 illustrates the procedure in detail. Both the sets of referents (i.e., with gesture choices and without gestures choices) and the order of the referents in each set was randomized to minimize any bias. We did this by asking participants to first pick a card from a set of cards lying on the desk in front of them, turn the card around, and then to read out loud the referent's description on the card. Consequently, the video recording(s), which matches the textual description of the referent, was presented on the tablet. Participants could watch the recording(s) multiple times, if needed.

In case gesture choices existed for a referent participants were asked to watch the gesture recording first and to choose the gesture that in their opinion matched best the corresponding effect. Then we asked participants to rate their gesture choice on two Likert scales, depicting *ease of use* and *goodness of mapping* of the gestures. We did this following the example provided by Wobbrock et al. [42]. Afterwards we asked participants explicitly if they could think of a new and different gesture, which fits the effect better than the proposed choices. If participants proposed a new gesture, we asked them again to rate their choice, considering ease of use and goodness of mapping of the gesture. Then participants were asked to explore the use of the gesture in each of the three arm positions and demonstrate how they would perform the gesture starting from the proposed arm positions. Participants were explicitly informed that for each arm position they could choose a different gesture without considering any (potential technical) feasibility issues. They were also asked to "think aloud" while considering different gestures.

4 RESULTS

We collected three different kinds of data: (i) all sessions were video recorded (image and audio), (ii) participants were asked to state their preferences if multiple choices were available, ratings of gestures (considering ease of use, goodness of mapping) were collected on Likert scales and (iii) notes were taken during each session by the second researcher focusing on mental models and metaphors applied or mentioned by the participants.

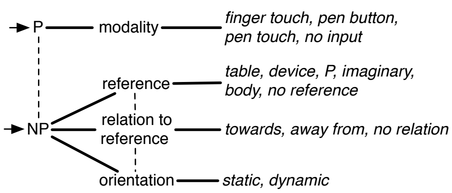

Figure 5: Gesture taxonomy extension tailored for restful bi-manual Pen + Mid-Air input on tablets.

4.1 Taxonomy of Gestures

Following the procedure demonstrated by Wobbrock et al. [42] two researchers manually classified all preferred mid-air gesture recordings along the dimensions: form, nature, binding, and flow. Similar to related work, which have adapted and extended these basic dimensions to their application and interaction contexts, we provide in Figure 5 a taxonomy extension tailored to restful bi-manual pen and mid-air input. In this extension each hand is classified separately. The focus for the non-preferred hand is on a *reference* "object" (e.g., or the users' own body, the preferred hand, or the table), which the hand *relates* to through its *orientation*, which can be completely static (e.g., palm is always oriented towards the table) or dynamic (e.g., palm is not always oriented towards the table but the table is used as a physical reference/anchor).

In contrast, the focus of the preferred hand is on input *modality* (i.e., pen touch, finger touch, pen button, and pen). Pen touch, finger touch and pen button were the main modalities proposed by participants. We have added the general term "pen" to classify any other input modality participants proposed with the pen. For example, one participant proposed to put the pen down on the table and pick it up again to trigger a command. While participants occasionally explored such "pen gestures" only once was a pen gesture preferred by a participant. We believe that one of the reasons for participants not preferring (complex) gestures/input with the pen was that many of the commands required participants to already touch the screen (and sometimes draw on the screen).

Figure 6 shows on the left side the classification results for all gestures and for all arm positions: non-rest (N), elbow rests on table (E), and arm rests on table (A). On the right side, Figure 6 shows agreement score for all referents and arm positions. Agreement scores reflect the degree of consensus between participants, considering proposed gestures for the same referent [42]. It also shows, for example, that for 5 of 13 referents; i.e., 3 (Rotate Canvas),4 (Change ink color of pen), 5 (Open/Close menu), 6 (Change brush), and 7 (Undo/Redo) aggreement scores were the same over all three arm positions. Furthermore, the aggreement scores for the $non-rest$ arm position is always lower or the same as the aggreement scores for the two rest positions $elbow\ rest$ and $arm\ rest$, but for referents 9 (Draw&change ink size) and 10 (Draw&change ink transparency) for which the agreement scores for the $arm\ rest$ positions are lower.

4.2 User Preferences and Observations

4.2.1 Choice-based Preferences. For the set of referents (with referent IDs 1 to 8) for which based on related work gesture choices were proposed, users preferred in 69% of cases one of the proposed gesture choices; that is, in 31% of cases participants still preferred a new gesture which they designed themselves over the available gesture choice(s). Left side of Figure 7 presents in detail user preferences for the set of gestures for which participants had gesture choices, differentiating preferences based on the three different arm positions. For example, for referents 7 (Undo/Redo) and 8 (Zoom) participants choose to propose new gestures more often for the *arm rested postion* (A) than for the arm positions *elbow rested* (E) and *non-rest* (N)

4.2.2 Self-reports of goodness of mapping and ease of use. Overall, we collected 471 ratings (i.e., 390 ratings for preferred gestures and 81 for non-preferred gestures) based on five-point Likert scales. Ratings for non-preferred gestures were collected either when a gesture choice was proposed but participants preferred an other gesture they designed themselves. Or, in a few cases (i.e., 13) when participants designed two different gestures for the same referent (and/or same arm position) and provided ratings for both, including a preference for one of them. Right side of Figure 7 provides an overview of participants ratings for *goodness of mapping* and *ease of use* for each preferred gesture differentiated by (i) the three arm positions referred to as modality and (ii) referents with and without gestures choices. Ratings for both, *goodness of mapping* and *ease of use* are high, assuring that participants seem overall satisfied with their gesture preferences. Right end side of Figure 7 provides an overview of ratings for gestures, which were not preferred by participants. As one would expect non-preferred gestures were rated lower at least at one of the two dimensions *goodness of mapping* and *ease of use*. During the study we have observed participants arguing that *ease of use* was more important than *goodness of mapping*, especially when one is already tired and rests their arm. This observation may explain why the mean ratings of *ease of use* are higher than the mean ratings for *goodness of mapping*.

A Pearson correlation test on the variables *goodness of mapping* and *ease of use* for the different groups, which are depicted on the right of figure 7 showed a correlation within the preferred gestures without gesture choices group ($r=0.46$, $p<0.001$). A less strong correlation was observed within the preferred gestures with gesture choices group ($r=0.28$); and no correlation within non-preferred gestures ($r=0.13$, $p=0.30$). No correlation between the two variables *goodness of mapping* and *ease of use* in the non-preferred gestures group means that participants' ratings shows no direct linear relationship between the two variables. Based on our experience, the reason is that participant occasionally rejected a gesture with a good mapping because its ease of use was not acceptable. For example, a paricipant argued for the *arm rested* position "the gesture is difficult to perform because of the movement restrictions"; another argued "actually, I would have to move my arm up to be able to perform the gesture". Yet another participant argued "there really isn't much much space free to move one's arm backwards" and explained how it is difficult to move ones arm forwards or backwards while it is rested on the table due to the arm's friction.

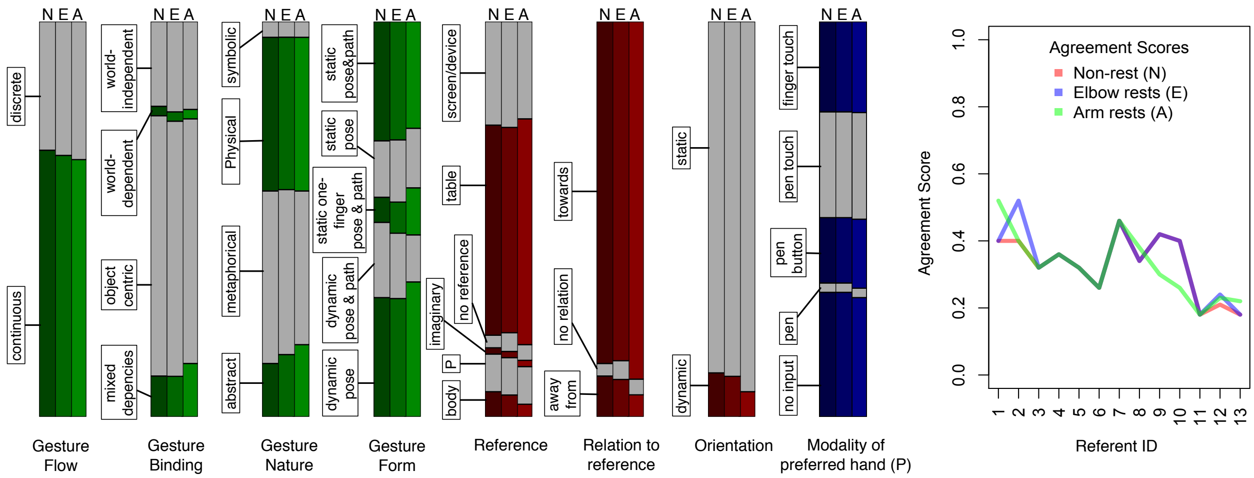

Figure 6: Categorisation of pen + mid-air gestures for the three arm positions: non-rest (N), elbow rested (E), and arm rested (A) is presented on the left hand side. The right hand sides provides the overview of agreement scores for all referents considering the three arm positions.

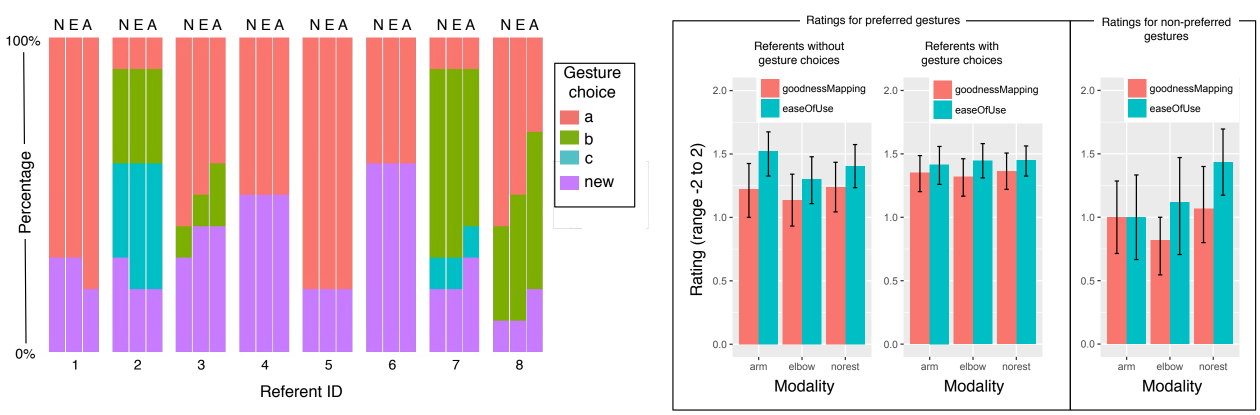

Figure 7: Summary of user preferences. On the left side an overview is presented for the "first" set of referents 1-8 (for which gesture choices were proposed) depicted for the three arm positions: non-rest (N), elbow rested (E), and arm rested (A). Note that Table 1 provides the relation between gesture choice and Referent ID. On the right hand side user ratings on *goodness of mapping* and *ease of use* are presented for the types of gestures(i.e., preferred and non-preferred gestures). Preferred gestures are further divided in referents with and without gesture choices. Error bars denote 95% (bootstrap) confidence intervals.

4.3 Resulting Gesture-Operation Mappings

The analysis of all preferred gestures (for all arm positions), which were either designed or chosen by participants for each referent was grouped in 38 different gestures. Following the procedure by Wobbrock et al. [41] agreement scores were utilized to identify final gestures for each of the 13 referents (see Figure 8). While participants often agreed/preferred to use the same gesture in all arm positions there were two exceptions for the *arm rest* position; i.e., for referents 8 (Zoom) participants preferred to use a *two finger pinch* instead of a *five finger grasp and move* and for referent 10 (Draw&change ink transparency) they preferred a *one/two finger linear movement* over a *flat hand linear movement*. Figure 9 provides exemplary gestures performed in rest positions.

Participants argued that whenever a command is associated with pen functions (e.g, change ink size or ink color of pen) the NP gesture should be used in parallel with pen input (e.g., pen touch or pen button). One participant stated *"I would want to do something on the pen to make clear that my gesture relates to the pen. If it [the pen] has a button I would press the button or if it [the pen] can recognize my grip I would grip harder"*. For example, while a <pinch gesture> + <pen touch> should change ink size, a <pinch gesture> + <finger touch> should zoom in or out of the canvas. Participants also argued that when opening a menu pushing the pen button should open a

Referent ID - Operation	NP gesture	P modality
1 - Pan: Move canvas	flat hand pan	finger touch/button
2 - Change ink size of pen	5 finger pinch	pen touch/button
3 - Rotate canvas	4 finger rotate	finger touch
4 - Change ink color of pen	make/open fist	pen touch/button
5 - Open/Close menu	flat hand push/shove	no input
6 - Change brush	make fist	pen touch/button
7 - Undo/Redo	flat hand linear movement	pen button
8 - Zoom	5 finger grasp and move	finger touch
Alternative	2 finger pinch	finger touch
9 - Draw&change ink size	5 finger pinch	pen touch/button
10 - Draw&change ink transp.	flat hand linear movement	pen touch/button
Alternative	1 finger linear movement	pen touch/button
11 - Switch to color picker	hand holds pipette	no input
12 - Draw&change ink color	1 or 2 finger cirlce	pen touch/button
13 - Switch to eraser	hand wiggles	no input

Figure 8: A final set of mappings of bi-manual input and operations based on agreement scores of participants. All conflicts are resolved by combining NP gestures with P input modalities (i.e., pen touch, pen button, and finger touch). Alternatives are for cases when users could not clearly agree on one specific gesture.

menu related with pen functions/input (e.g., for choosing a color or a different brush). All NP gesture conflicts could be resolved by combining input modality for P with the agreement scores. In cases the agreement scores did not clearly favor one gesture the second gesture is provided (see Figure 8) as an alternative gesture.

4.4 Contextual Gesture Differences

In Figure 8 we presented the final (minimal) set of mappings of bi-manual "gestures" and procedures across all three contexts of interaction that participants agreed on. This set of (bi-manual) gestures is based on commonalities across users and contexts. However, (and please note) that there are still differences in how the same gesture is typically executed in each context. When users take restful arm positions, designers/developers should expect arm/hand gestures do become smaller with more use of wrist movements (including hand rotations), and single finger movements. Furthermore, typical movement directions and trajectories change depending on the specific rest position. While in a non-rest position arm movement trajectories tend to be straight in rest positions movements show more curvature.

In Figure 10 we depict contextual differences of the interaction space, considering all the 38 gestures proposed/preferred by participants (including gestures that multiple participants did not agree on) and on participants general gesture exploration behavior throughout the study (including gestures explored but not preferred by a participant). We identified differences in (i) shape and size of the mid-air space explored for interaction, (ii) main arm movement directions, and (iii) exploration tendencies in hand/finger movements.

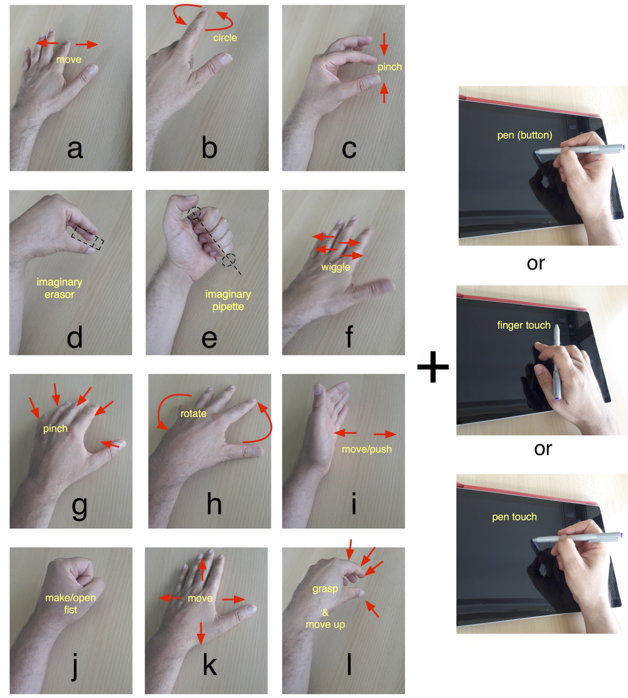

Figure 9: Exemplary visualization of restful gestures with NP as explored and proposed by participants; and a visualization of the main input modalities for P, which can be used in combination with NP gestures.

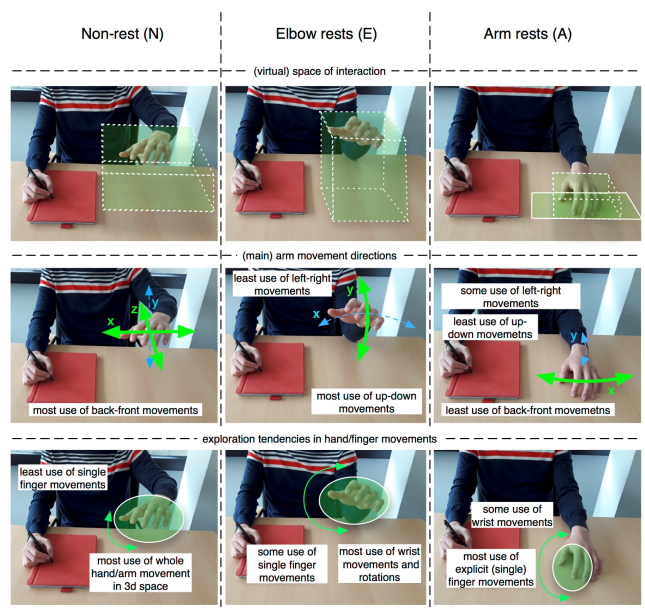

Figure 10: Contextual differences in gestural interaction.

5 DISCUSSION

We have argued in the beginning that allowing users to rest their arm while performing hand gestures would limit the inherent effects of gestural mid-air interfaces and thus prevent fatigue and discomfort. Our intention was not to study how users would (want to) perform gestures when already fatigue but how users would want to perform gestures in alternative "contexts" in which their arms are rested to prevent arm fatigue or prolong the time until fatique may hit. The research question was about how users would perform gestures in (alternative) rest positions compared to non-rest positions, and even more, if they would prefer the same gestures in rest positions or prefer different gestures for different arm positions?

We have found that for many operations and gestures participant would want the same gestures in each context but due to contextual (e.g., physical) constraints have to adapt the way gestures are performed. We observed, for example, that gestures will change in size and shape, and wrist and finger movements will replace larger arm and hand movements in 3D space.

Elicitation studies aim (in general) to produce easy to use interfaces, addressing mainly requirements of novice users. Since, agreements scores put emphasis on commonalities in gestures, results are often simple hand postures and movements. For example, for referent 13 (change to eraser tool) many participants proposed to wiggle their NP but a similar amount of participants proposed to use an imaginary eraser (see Figure 9 f); however, we observed individual difference for how participants imagined they would hold and use an imaginary eraser, which were a challenge for the existing categorization schemes for gestural interaction. Furthermore, we have (arguably) observed a legacy bias because all participants were regular touch device users and already knew about touch gestures. However, we did expect and welcome prior touch gesture experience, since we envision a multimodal interaction space that combines and uses touch and mid-air gestures interchangeably.

In order to categorize gestures performed by NP during bimanual interaction we saw potential to extend and modify the existing gesture categorizations provided by previous work (i.e., [3, 41]). We extended the categorization with three dimensions (i.e., Reference object, Relation to reference object, and NP orientation) explaining the relationship between NP and a reference object used during mid-air gestures. Considering the input modality for P our results are not different form results reported in previous work [3]. Participants argued that whenever the mid-air gesture influences pen ink (or something related to the pen) than P (i.e., pen button) should be pressed. For example, one participant stated "[I would want to] press the pen [button] because I want to change something about the pen". Another participant stated "[I would] press pen, because I want to change the pen color". Most participants did not differentiate between finger touch and pen touch. Some participants argued they would like to also use pen pressure on the tablet.

Especially in the *arm rested* position participants naturally made use of their touch senses, which resulted in an exploration of smaller and finger-based gestures. Often participants argued that they would want to use a "typical touch gesture", which they already learned to use and accepted for similar use contexts. Sometimes they would perform the gestures very similar to a touch gesture on the table but at times they would perform the gesture mid-air. For example, most participants preferred a two finger pinch over a five finger pinch when performed on the table. Participants stated, for example "[this hand/finger movement] reminds me of classical tablet movements", "mapping is very good because I already know it from a different context", and "it [the gesture] is more intuitive because one is accustomed to it".

All but one participants stated that in the *arm rested* position moving the arm was harder and required more force due to a large friction surface on the table. Participants argued that they either would move their arms slightly up (i.e., leaving the rest position for a short time to perform quickly the gestures and move back to the rest position) or prefer to perform an equivalent version of the gesture with their fingers. One participant argued for the arm rest position "I would not perform the gestures in [3D] space but horizontally [on the surface of the table]". While a few participants argued that it would be beneficial if the table could recognize touch, most have used the touch sensation only to modulate their finger movement but did not seem to assume touch was the interaction modality. Depending on the workplace, one could consider addressing the the issue with the large friction surface when the arm is rested on the desktop by wearing an arm sleeves (made of smooth material) but it would not a general solution.

Overall, results should be regarded carefully considering the narrow sample of participants and the context of the presented research. Participants were aware of fatigue as an issue in mid-air gestural interaction, which may have influenced their preferences, resulting in preferring easier to use gestures over gestures with better mappings. Concern for fatigue may also have limited participants tendency to explore more complex gestures with (potentially) better mappings for restful arm positions. One participant, for example stated "if I have already taken this position [arm rested on desk] I wouldn't want to move much".

The work at hand contributes (in general) to the ongoing study of the multi-modal design space for input on interactive screens and surfaces. We hope that the concrete set of gesture-operation mappings which we elicited for drawing related applications will help and inspire researchers and practitioners who develop gesture-based interfaces. Furthermore, we also hope that the research model we described in this paper is apt for replication for different input modalities and fellow researchers will replicate it whenever elicitation of contextual user preferences are critical.

6 CONCLUSION

We reported on a user and choice based elicitation study, exploring contextual preferences and alternatives for pen + mid-air gestural input on tablets. A detailed analysis of user preferences and behavior based on self-reports and a categorization of gestures based on video recordings was provided. Moreover, commonalities and differences were highlighted in gesture choices of users and how gestures are executed when users perform mid-air gestures in alternative use contexts (e.g., elbow rested on table, arm rested on table) between which users may alternate to prevent arm fatigue.

We hope the insights that we provided considering multiple dimensions will inspire and guide future designers/developers in anticipating how gestures will vary over different contexts and how to utilize gesture recognizers to enable restful gestural interaction.

REFERENCES

[1] Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2011. Medusa: A Proximity-aware Multi-touch Tabletop. In *Proc. of (UIST '11)*. ACM, 337–346.

[2] Ilhan Aslan and Elisabeth André. 2017. Pre-touch Proxemics: Moving the Design Space of Touch Targets from Still Graphics Towards Proxemic Behaviors. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI 2017)*. ACM, New York, NY, USA, 101–109. https://doi.org/10.1145/3136755.3136808

[3] Ilhan Aslan, Ida Buchwald, Philipp Koytek, and Elisabeth André. 2016. Pen + Mid-Air: An Exploration of Mid-Air Gestures to Complement Pen Input on Tablets. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 1, 10 pages. https://doi.org/10.1145/2971485.2971511

[4] Ilhan Aslan, Alina Krischkowsky, Alexander Meschtscherjakov, Martin Wuchse, and Manfred Tscheligi. 2015. A Leap for Touch: Proximity Sensitive Touch Targets in Cars. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. ACM, New York, NY, USA, 39–46. https://doi.org/10.1145/2799250.2799273

[5] Ilhan Aslan, Thomas Meneweger, Verena Fuchsberger, and Manfred Tscheligi. 2015. Sharing Touch Interfaces: Proximity-Sensitive Touch Targets for Tablet-Mediated Collaboration. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, NY, USA, 279–286. https://doi.org/10.1145/2818346.2820740

[6] Ilhan Aslan, Feiyu Xu, Hans Uszkoreit, Antonio Krüger, and Jörg Steffen. 2005. COMPASS2008: Multimodal, Multilingual and Crosslingual Interaction for Mobile Tourist Guide Applications. In *Intelligent Technologies for Interactive Entertainment*, Mark Maybury, Oliviero Stock, and Wolfgang Wahlster (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 3–12.

[7] Myroslav Bachynskyi, Gregorio Palmas, Antti Oulasvirta, Jürgen Steimle, and Tino Weinkauf. 2015. Performance and Ergonomics of Touch Surfaces: A Comparative Study Using Biomechanical Simulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1817–1826. https://doi.org/10.1145/2702123.2702607

[8] Myroslav Bachynskyi, Gregorio Palmas, Antti Oulasvirta, and Tino Weinkauf. 2015. Informing the Design of Novel Input Methods with Muscle Coactivation Clustering. *ACM Trans. Comput.-Hum. Interact.* 21, 6, Article 30 (Jan. 2015), 25 pages. https://doi.org/10.1145/2687921

[9] Lonni Besançon, Mehdi Ammi, and Tobias Isenberg. 2017. Pressure-Based Gain Factor Control for Mobile 3D Interaction Using Locally-Coupled Devices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1831–1842. https://doi.org/10.1145/3025453.3025890

[10] L. BesanÃğon, P. Issartel, M. Ammi, and T. Isenberg. 2017. Hybrid Tactile/Tangible Interaction for 3D Data Exploration. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan 2017), 881–890. https://doi.org/10.1109/TVCG.2016.2599217

[11] Peter Brandl, Clifton Forlines, Daniel Wigdor, Michael Haller, and Chia Shen. 2008. Combining and Measuring the Benefits of Bimanual Pen and Direct-touch Interaction on Horizontal Interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '08)*. ACM, New York, NY, USA, 154–161. https://doi.org/10.1145/1385569.1385595

[12] Xiang 'Anthony' Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff, and Scott E. Hudson. 2014. Air+Touch: Interweaving Touch & In-air Gestures. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 519–525. https://doi.org/10.1145/2642918.2647392

[13] Christian Corsten, Simon Voelker, Andreas Link, and Jan Borchers. 2018. Use the Force Picker, Luke: Space-Efficient Value Input on Force-Sensitive Mobile Touchscreens. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 661, 12 pages. https://doi.org/10.1145/3173574.3174235

[14] Nem Khan Dim, Chaklam Silpasuwanchai, Sayan Sarcar, and Xiangshi Ren. 2016. Designing Mid-Air TV Gestures for Blind People Using User- and Choice-Based Elicitation Approaches. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 204–214. https://doi.org/10.1145/2901790.2901834

[15] ROGER M Enoka and DOUGLAS G Stuart. 1992. Neurobiology of muscle fatigue. *Journal of applied physiology* 72, 5 (1992), 1631–1648.

[16] Dustin Freeman, Ramadevi Vennelakanti, and Sriganesh Madhvanath. 2012. Freehand pose-based Gestural Interaction: Studies and implications for interface design. In *Proc. of IHCI'12*. IEEE, 1–6.

[17] Mathias Frisch, Ricardo Langner, and Raimund Dachselt. 2011. Neat: A Set of Flexible Tools and Gestures for Layout Tasks on Interactive Displays. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/2076354.2076356

[18] Darren Guinness, Alvin Jude, G. Michael Poor, and Ashley Dover. 2015. Models for Rested Touchless Gestural Interaction. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 34–43. https://doi.org/10.1145/2788940.2788948

[19] Jože Guna, Grega Jakus, Matevž Pogačnik, Sašo Tomažič, and Jaka Sodnik. 2014. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors* 14, 2 (2014), 3702–3720.

[20] Otmar Hilliges, Shahram Izadi, Andrew D. Wilson, Steve Hodges, Armando Garcia-Mendoza, and Andreas Butz. 2009. Interactions in the Air: Adding Further Depth to Interactive Tabletops. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology (UIST '09)*. ACM, New York, NY, USA, 139–148. https://doi.org/10.1145/1622176.1622203

[21] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[22] Ken Hinckley, Seongkook Heo, Michel Pahud, Christian Holz, Hrvoje Benko, Abigail Sellen, Richard Banks, Kenton O'Hara, Gavin Smyth, and William Buxton. 2016. Pre-Touch Sensing for Mobile Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2869–2881. https://doi.org/10.1145/2858036.2858095

[23] Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddington, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. 2010. Pen + Touch = New Tools. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 27–36. https://doi.org/10.1145/1866029.1866036

[24] Felix Kistler and Elisabeth André. 2013. User-Defined Body Gestures for an Interactive Storytelling Scenario. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 264–281.

[25] Nicolai Marquardt, Ricardo Jota, Saul Greenberg, and Joaquim A. Jorge. 2011. The Continuous Interaction Space: Interaction Techniques Unifying Touch and Gesture on and above a Digital Surface. In *Human-Computer Interaction – INTERACT 2011*, Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 461–476.

[26] Fabrice Matulic and Moira C. Norrie. 2013. Pen and Touch Gestural Environment for Document Editing on Interactive Tabletops. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS '13)*. ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2512349.2512802

[27] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. FingerIO: Using Active Sonar for Fine-Grained Finger Tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1515–1525. https://doi.org/10.1145/2858036.2858580

[28] Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum. 2004. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction*, Antonio Camurri and Gualtiero Volpe (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 409–420.

[29] Donald A Norman and Stephen W Draper. 1986. User centered system design. *Hillsdale, NJ* (1986), 1–2.

[30] Mohammad Obaid, Markus Häring, Felix Kistler, René Bühling, and Elisabeth André. 2012. User-Defined Body Gestures for Navigational Control of a Humanoid Robot. In *Social Robotics*, Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, and Mary-Anne Williams (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 367–377.

[31] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined Gestures for Augmented Reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 955–960. https://doi.org/10.1145/2468356.2468527

[32] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 197–206. https://doi.org/10.1145/1978942.1978971

[33] Chaklam Silpasuwanchai and Xiangshi Ren. 2015. Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies* 80 (2015), 1–13.

[34] Giovanni Maria Troiano, Esben Warming Pedersen, and Kasper Hornbæk. 2014. User-defined Gestures for Elastic, Deformable Displays. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/2598153.2598184

[35] Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y. Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3327–3336. https://doi.org/10.1145/2702123.2702214

[36] Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proceedings of the 10th European Conference on Interactive TV and Video (EuroITV '12)*. ACM, New York, NY, USA, 45–48. https://doi.org/10.1145/2325616.2325626

[37] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. 2011. Vision-based Hand-gesture Applications. *Commun. ACM* 54, 2 (Feb. 2011), 60–71. https://doi.org/10.1145/1897816.1897838

[38] Andrew D. Wilson, Shahram Izadi, Otmar Hilliges, Armando Garcia-Mendoza, and David Kirk. 2008. Bringing Physics to the Surface. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08)*. ACM, New York, NY, USA, 67–76. https://doi.org/10.1145/1449715.1449728

[39] Markus L. Wittorf and Mikkel R. Jakobsen. 2016. Eliciting Mid-Air Gestures for Wall-Display Interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 3, 4 pages. https://doi.org/10.1145/2971485.2971503

[40] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

[41] Jacob O. Wobbrock and Krzysztof Z. Gajos. 2008. Goal Crossing with Mice and Trackballs for People with Motor Impairments: Performance, Submovements, and Design Directions. *ACM Trans. Access. Comput.* 1, 1 (May 2008), 1–37. https://doi.org/10.1145/1361203.1361207

[42] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[43] Mike Wu, Chia Shen, Kathy Ryall, Clifton Forlines, and Ravin Balakrishnan. 2006. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *Horizontal Interactive Human-Computer Systems, 2006. TableTop 2006.* IEEE, 8–pp.

[44] Xing-Dong Yang, Tovi Grossman, Pourang Irani, and George Fitzmaurice. 2011. TouchCuts and TouchZoom: Enhanced Target Selection for Touch Displays Using Finger Proximity Sensing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2585–2594. https://doi.org/10.1145/1978942.1979319

[45] Dongwook Yoon, Nicholas Chen, and François Guimbretière. 2013. TextTearing: Opening White Space for Digital Ink Annotation. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 107–112. https://doi.org/10.1145/2501988.2502036

[46] Chen Zhao, Ke-Yu Chen, Md Tanvir Islam Aumi, Shwetak Patel, and Matthew S. Reynolds. 2014. SideSwipe: Detecting In-air Gestures Around Mobile Devices Using Actual GSM Signal. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 527–534. https://doi.org/10.1145/2642918.2647380