On Gesture Combination: An Exploration of a Solution to Augment Gesture Interaction

William Delamare[1†], **Chaklam Silpasuwanchai**[1,2], **Sayan Sarcar**[1,3]
Toshiaki Shiraki[1], **Xiangshi Ren**[1*]
[1]Kochi University of Technology, Japan, [2]Asian Institute of Technology, Thailand, [3]University of Tsukuba, Japan
{william.delamare, xsren}@acm.org, {chaklam.silpasuwanchai, mailtosayan, toshiraki}@gmail.com

ABSTRACT
Current gesture interaction paradigm mainly involves a one-to-one gesture-command mapping. This leads to memorability issues regarding (1) the mapping - as each new command requires a new gesture, and (2) the gestures specifics (e.g., motion paths) - that can be complex to leverage the recognition of several gestures. We explore the concept of combining 3D gestures when interacting in smart environments. We first propose a design space to characterize the *temporal* and *spatial* combination aspects, and the *gesture types* used by the combination. We then report results from three user studies in the context of smart TV interaction. The first study reveals that end-users can create gesture sets with combinations fully optimized to reuse gestures. The second study shows that combining gestures can lead to improved memorability compared to single gestures. The third study reveals that preferences for gestures combination appear when single gestures have an abstract gesture-command mapping.

Author Keywords
Two-handed interaction; freehand gestures;
mid-air interaction; gestural interaction; gesture combination.

CCS Concepts
•**Human-centered computing** → **Gestural input;**

INTRODUCTION
Mid-air hand gestures have gained much attention to interact with augmented physical appliances [15, 17, 40]. Indeed, gesture interaction has the potential to allow users to control their surroundings from a distance. Thus, researchers proposed solutions to improve gesture recognition [10, 13], gesture registration [41], gesture set designs [33, 35], gesture guidance [8, 27] or toward easing the deployment [38] and the adoption [31] of gesture interaction to name a few.

Past studies mainly involve actions and movements of one hand, or sometimes coordination among two hands [17, 24] to perform a single gesture-command mapping. Each one or two-handed gesture corresponds to a specific command. However, as the number of commands increases, it becomes increasingly difficult for designers to create distinguishable gestures, and for users to memorize the gesture mappings.

We propose to explore the concept of *gestures combination* [20]. For instance, a first gesture - a signifier gesture - could signify a context (e.g., 'Light') while a second gesture - an operator gesture - could trigger a command (e.g., 'Switch on'). The expected theoretical main benefit of such gestures combination resides in the re-usability of gesture definitions. For instance, users can memorize the 'increase/decrease' command and use it for a light brightness, a sound volume, a temperature setting, etc... Gesture combinations can easily extend a set of commands without increasing the number of gestures to memorize. However, previous work explored only sequences of 2D stroke gestures [18, 20]. We hence propose to expand this investigation to 3D mid-air gestures combinations.

We first propose a design space to characterize the *temporal* and *spatial* aspects of the combination, along with the *gesture type* used by the combination. We then propose a series of studies to explore the use of gestures combinations in the context of smart TV interaction. A focus group workshop revealed that participants mainly focused on relaxed sequential combinations, without focusing on spatial aspects. In addition, participants proposed to use static gestures with the non-dominant hand to specify the context, and static or dynamic gestures with the dominant hand for the operator.

Doing so, participants fully optimized the reuse of gestures in the designed gesture set. We next compare the memorability of the previously defined gesture sets (single gestures and combinations). Results show that combining gestures is a valid option to prevent memorability problems due to unclear single gesture/command mappings. Lastly, we assess users' preferences between single gestures and gestures combination. Results reveal that users prefer gestures combination when single gestures counterparts have an unclear gesture/command mapping, confirming the benefits of combining gestures.

The contributions of our paper are: (1) a design space which

[†]JSPS fellow
[*]Corresponding author

ISS '19 November 10–13, 2019, Daejeon, Republic of Korea

DOI: https://doi.org/10.1145/3343055.3359706

can guide the development of novel gesture combinations; and (2) an investigation into the design (study 1), the memorability (study 2), and users' preferences (study 3) regarding such gesture combinations.

RELATED WORK

We present previous works regarding smart environment interaction with an emphasis on gesture interaction, gestures combination, and results from studying two-handed interaction.

Smart Environment Gesture Interaction

Gesture interaction in smart environments [15, 17, 26] mostly involve semaphoric dynamic gestures [1, 14]. Semaphoric gestures define a vocabulary to trigger discrete commands. Dynamic gestures involve motions. For instance, users can draw a circle in mid-air to trigger a 'Record' command [17].

Among appliances in smart environments, several studies explored TV interaction [11,17,37]. Indeed, gesture interaction holds the potential to be a good complement to the conventional remote control. In addition, TV is an interactive object which proposes several features (brightness, volume, etc) and several commands (up, next, play, etc). This makes it a good candidate to explore interaction paradigm such as voice recognition [25] or freehand gesture interaction [32, 34, 36, 37].

Most previous works report the use of one-to-one gesture/command mappings. In other words, it demands designers to come up with a new unique gesture for every command. For example, although "Increase Volume" & "Increase Brightness" may intuitively share similar gesture preferences (e.g., "move the hand upward" [17]), designers have to create different gestures for each command, distinct enough to enhance gesture recognition performance. This can lead to a heavy memory load for users. We hence want to explore the possibilities offered by combining gestures.

Gestures Combination

Combining gestures allows to divide the sets of available commands into hierarchy sets [7, 16, 18, 20]. For instance, instead of having 36 items to trigger, users could perform one gesture to select one command among six in a first level, and one gesture to select one command among six in a second level. Hierarchies also allow to reuse gestures since a context is specified in the hierarchy definition. For instance, with Hierarchical Marking Menu, a left stroke can be executed in all levels and have different meanings according to the previously selected levels [18]. In addition, hierarchical structures can improve learnability [20]. Indeed, hierarchical structures can benefit from the similarity between commands by reusing common gestures. For instance, once 'Light' or 'Kettle' is selected, a common 'on/off' gesture could be executed.

However, to the best of our knowledge, very little work took interest in combining gestures, and only focused on *sequential* gestures combinations, in 2D [18, 20, 28, 29] or 3D [7, 16]. We argue that the design space is larger than what has been explored. For this, we extend the concept of sequential gestures (i.e. "chaining primitives" [20]) to the broader notion of *gesture combination*.

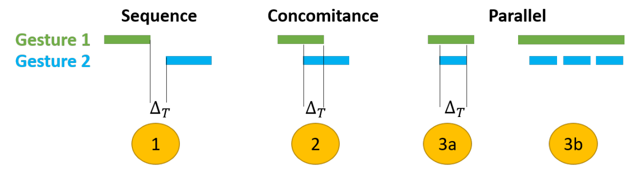

Figure 1. Illustration of temporal relationships between two gestures. 1) Sequence. 2) Concomitance. 3a) Parallel combination of two gestures. 3b) Parallel combination of one gesture and a sequence of gestures. Δ_T represents a delay specific to each combination, which can be used to transmit additional information in a *strict* temporal relationship.

Two-Handed Interaction

Buxton and Myers found that users can effectively use both hands concurrently to enhance performance over one-handed operations [5]. Some work suggested that symmetric two-handed interaction (both hands doing the same role, e.g., typing) can be used effectively in a number of tasks such as 2D or 3D navigation [2, 3].

Guiard's kinematic chain theory [12], the best-known model of two-handed interactions, stated that having the non-dominant hand serves as a coarse frame of reference (e.g., holding a paper) while the dominant hand performs detailed work on that frame (e.g., writing on the paper) is the most effective combination. This model has since been proven and exploited by many other works. For example, zoom and pan interaction (i.e., the dominant hand points to a spatial reference while the non-dominant hand controls the zoom action) [19, 23], rotation and scaling interaction (uses similar method as zoom) [4], and game interaction (shooting while using first-aid kit) [30] or 3D virtual objects manipulation [24] have been explored.

Summary

The literature review reveals that (i) gesture input is a good candidate for smart environment interaction, but current gesture sets mainly involve one-to-one gesture/command mappings, (ii) gestures combination can lead to improve learnability thanks to gesture reuse, but only explored sequences of 2D gestures so far, and (iii) two-handed gesture interaction can enhance input performance. Similarly, we acknowledge the two-handed capabilities and aim to explore the design space offered by combining two-handed gestures.

DESIGN SPACE

Our design space considers both the *combination* aspect and the *gestures* aspect (Figure 2). We examine two features of the combination: the temporal and spatial relationships [39].

Combination Aspect: Temporal Relationship

We define the temporal *nature* and *rigidity* of a combination.

Nature

A **sequential** combination involves a temporal gap between gestures execution (Figure 1, 1). For instance, a user can execute a gesture to select 'Light' followed by a gesture to trigger the 'On' command. A sequential combination allows users to execute the gestures with one or two hands. A **parallel** combination involves the execution of a gesture *during* the execution of another one (Figure 1, 3a and 3b). This combination provides a more time-efficient way to input commands than

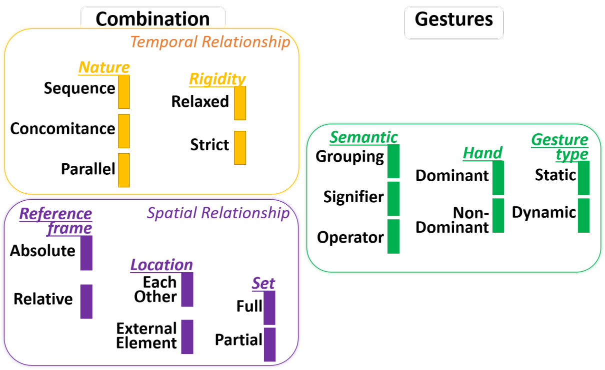

Figure 2. Illustration of the design space characterizing the *combination* aspect with two features: the temporal and spatial relationships, and the *gesture* aspect.

a sequential input. For instance, a user can execute a gesture to select 'Light' while also executing the gesture to trigger the 'On' command (Figure 1, 3a). We can also envision a user executing the gesture to trigger the 'On' command with one hand, and triggering 'Light', 'Music' and 'Kettle' with the other hand (Figure 1, 3b). In between sequential and parallel combinations exists a **concomitant** combination, which involves a time overlap of gestures execution (Figure 1, 2).

Rigidity

With a **relaxed** rigidity, the temporal relationship does not hold any information. This gives users peace of mind while executing the gestures free from any time constraint. On the contrary, with a **strict** rigidity, the relationship provides information via delays and/or duration. This gives users the possibility to input additional information (e.g. a parameter) at the cost of additional time constraints.

With a strict sequential combination for instance, the gap(s) between the gestures executions (e.g., Δ_T in Figure 1) could hold an additional meaning. For instance, the delay between 'Light' and 'On' could define the light intensity for a dimmable light bulb: the longer the gap, the brighter the intensity. With a strict parallel or concomitant combination for instance, if the 'On' gesture lasts X% as long as the 'Light' command, the intensity can be set to X%.

The nature and the rigidity of the temporal relationship are particularly relevant for three reasons. First, they characterize how comfortable users can be while triggering a command, such as if the execution requires one or two hands, and/or if it requires additional attention regarding the time component of the execution. Second, they characterize the theoretical time efficiency of the gesture input. For instance, a parallel combination is theoretically faster to execute than a sequential combination. Third, it characterizes the possibility to transmit additional information, hence increasing the input bandwidth of a standard gesture interaction.

Combination Aspect: Spatial Relationship

We define a spatial relationship of a gestures combination according to the gestures execution reference frame.

The *reference frame* can be **absolute**. In this case, the gestures execution locations do not transmit any additional meaning to the command. This is a desired property to give freedom to users as to where to execute the gestures. The reference frame can be **relative** and hence transmits additional information via the execution locations. The execution *locations* can be **relative to each other**. For instance, the distance between the two hands triggering the 'Light' and 'On' commands can specify the light intensity. In the event of a one-handed gesture interaction, the system can consider the distance between the two execution locations. The reference frame can also be **relative to an external element**. For instance, the distance between the couch and the 'Music' gesture and/or the 'On' gesture execution could specify the sound volume. In this case, we further describe the relative relationship whether it concerns the **full** *set* of gestures, or only a **partial** subset.

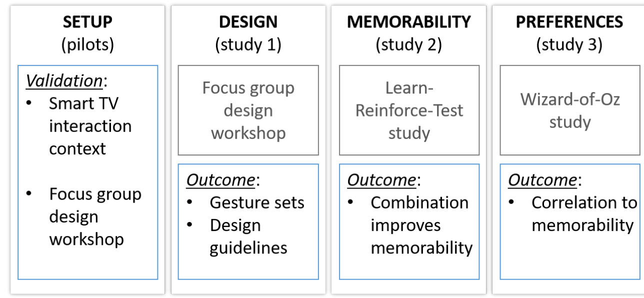

Figure 3. User experiments overview.

The reference frame of the spatial relationship is particularly relevant for two reasons. First, it can characterize the overall footprint (i.e. movement length) [6], and hence the potential fatigue due to the gesture execution. If we consider the same gesture shapes - one instance to be executed in any location, and one instance in a specific location - the overall footprint from the *intention* to the *beginning* of the gesture execution is potentially larger in the case of a relative spatial relationship due to extra motion required to reach the specific starting locations. Second, as for the temporal relationship, the spatial relationship characterizes the possibility to increase the input bandwidth of the standard gesture interaction.

Gesture Aspect

We consider three descriptive features. The **semantic** feature describes which part of the command the actual gesture refers to. For instance, in the sequence 'Living room' > 'Light' > 'On/Off', the gesture can refer to a grouping (e.g. 'Living room'), a signifier (e.g., 'Light') or an operator (e.g., 'On/Off'). We also consider the **hand** used for executing the gesture: dominant or non-dominant. These features can capture the relationship between the semantic and the hand used to trigger the command part. Lastly, we consider the **gesture type**: static (e.g., thumb up) or dynamic (e.g. circle motion path).

PILOT STUDIES

Prior to the three studies, we conducted two pilot studies (Figure 3). The pilot studies explored the suitability of (i) the TV interaction context, and (ii) a user-elicitation approach to explore the design space of gestures combination.

TV Interaction Context: Pilot Interview Study

We decided to run an interview with 14 participants in this early exploration stage for two reasons. First, it allowed us to get early feedback about combination of gestures via a concrete use case. Second, we wanted to validate the use of TV interaction to explore the concept of combination of gestures in terms of usefulness and usability.

Overall, participants would use gesture interaction if gestures are easy to remember and to execute. Four participants reported that they preferred to use gestures only for some frequently-used commands such as "Next Channel" or "Volume Up/Down", commenting that these gestures are easy to remember and thus easy to use. For other commands, participants preferred using the conventional remote control as they did not want to memorize new gestures. They also reported that using remote controllers was already part of their habit. However, most participants (10/14) reported that they preferred gestures over the remote control if the recognition system can work accurately and seamlessly. Participants reported that gestures are fun, easy to use, and can be used with another person in the room. They also reported that it is often a hassle to find the standard remote control.

Overall, the interview supported the idea of gestural TV interaction. This validates the fact that the TV interaction context (i) is a suitable context to explore the gesture interaction area, and that (ii) it admits the expected central limitation of standard gesture interaction, i.e. the increase of the number of gestures to remember with the increase of available commands. We hence consider only the TV interaction context in the remainder of this work. It allows us to focus on combinations of two gestures only. We discuss combinations of more than two gestures as an extension of the current work.

User-Elicitation Approach: Pilot Elicitation Study

Our design space greatly extends what end-users expect from a gesture interaction context. It is not clear whether the common approach for defining gestural interactions, i.e. user-elicitation studies, will be suitable for our purpose [42]. We hence conducted two user-elicitation studies. In the first study, we asked 15 participants to separately define 23 signifier-operator gesture combinations for a selected set of commonly used TV commands [37]. Because of low agreement scores, we then invited another 15 participants with a modified experimental protocol: participants had to define the signifiers and the operators separately to cover the same set of commands.

As expected, we found that an elicitation study is not suitable for exploring gestures combinations. When using Wobbrock et al.'s agreement score calculation [42], the agreement score of the combination designs in the first experiment was low (M = 0.11) for all commands. In the second experiment, the participants reported that many of the individually defined signifiers and operators were awkward to map into gestures when used jointly for a command. Thus, even without introducing the concept of *temporal* and *spatial* relationships, the agreement scores and feedback reflect a wide difference regarding how end-users would define gestures combinations.

We hence decided to organize a focus group design workshop instead of a user-elicitation study. We also decided to not introduce the design space. The rationale behind this decision was to avoid getting participants overwhelmed by the large number of design possibilities, while still being able to gather data to understand which features would be straightforward and clearly apparent to end-users, and which features would open new HCI research avenues for further investigations.

STUDY 1: DESIGN WORKSHOP

Our design workshop aimed to reveal what would be the main features of gestures combination participants would exploit, and to collect initial end-user feedback about this new concept. We conducted a semi-structured design workshop in order to leverage the group dynamics.

Participants

Eight university students and staff (one female, all right-handed, age 20 to 35, M = 29.25) volunteered for the study. All had knowledge about free-hand gestures (e.g., Kinect). Three reported having experiences with smart TVs.

Procedure

The design workshop consisted of four sessions. Two experimenters served as moderators. A third experimenter recorded the entire workshop and took notes on the side.

During the first session, we explained the background and motivations of the workshop to participants. We then showed videos of existing smart TV systems, their functions, and current standard gestural interactions to get all participants on the same understanding. We then introduced the concept of gestures combination. As an example, we asked participants to come up with one signifier and one operator gesture for "Increase Brightness" to help showcase the concept.

During the second session (1h), we led participants to freely share their opinions on three open-ended questions: 1) Do you think that a gestures combination can achieve the goal it claims (i.e., gesture re-usability and memorability)? 2) What are the pros and cons of combining gesture? 3) What are some key design considerations? At the end of the session, moderators summarized the design considerations emerged in the discussion and reviewed them with participants.

The third session (1h) intended to 1) identify commands likely compatible with a gestures combination; 2) identify gestures that could be reused by different commands; and 3) develop a gesture set based on the results of 1) and 2). The gestures design was a group effort: all participants could illustrate and share their opinions and ideas on a white-board or blank sheets of paper. We instructed participants to follow the design considerations from the second session. The moderators were not involved in the group design process to avoid biases. After participants reached a consensus, they presented their gesture set and the rationale behind it.

In the fourth session (30min), we asked participants to define the standard single gestures for commands raised in the third session but not covered in previous work (e.g., [37]). We used the two resulting gesture sets for further comparative evaluation in Studies 2 and 3.

Findings

We present the design workshop outcomes in two parts: the created gesture sets, and participants' feedback and design considerations. User feedback and design considerations were

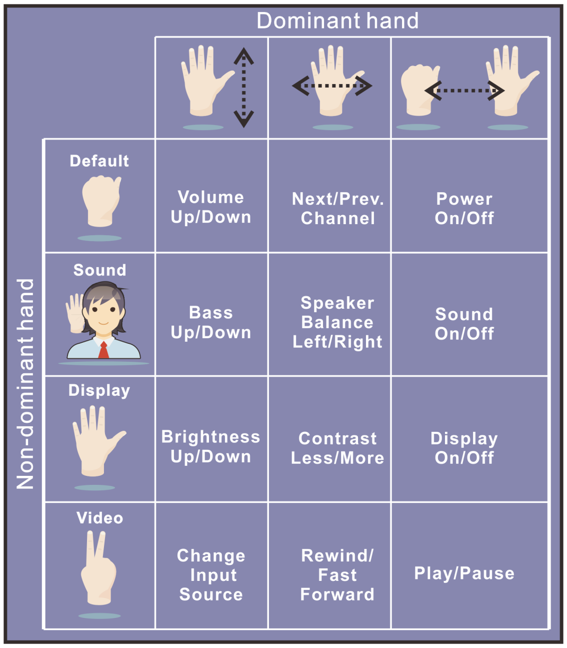

Figure 4. Gesture set while combining gestures. Gestures are depicted with rows showing the signifiers and columns showing the operators.

coded by three moderators. We used an open coding process to identify themes on user perception about combining gestures. We assigned codes to each of participants' responses. As we analyzed more responses, new codes emerged and existing ones were modified. This process allowed us to identify consistent themes among participants.

Gesture Sets

Figure 4 shows the gesture set when combining gestures. Participants picked out 12 pairs of compatible commands. Three commands ("Power on/off", "Volume up/down", "Next/Previous channel") were paired with the "Default" signifier containing the commonly-used commands. These "Default" commands can be performed without the signifier to allow quick access (Figure 4, first row). Some commands are rather specific to smart TV functions which were not defined in past work [37], such as "Bass+/-" and "Display on/off" when streaming music for instance.

Interestingly, participants managed to fully optimize the reuse of operator gestures across signifiers. Figure 5 illustrates the number of gestures to define in a context in which each signifier (e.g., augmented physical appliances) comes with three commands. In this case, the number of gestures in the standard

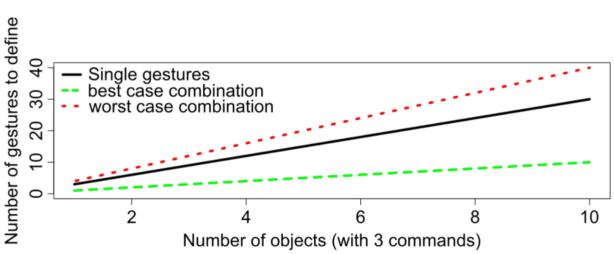

Figure 5. Theoretical illustration of the evolution of the number of gestures to define according to the number of signifiers (e.g., augmented physical objects) in the scenario with three commands per objects.

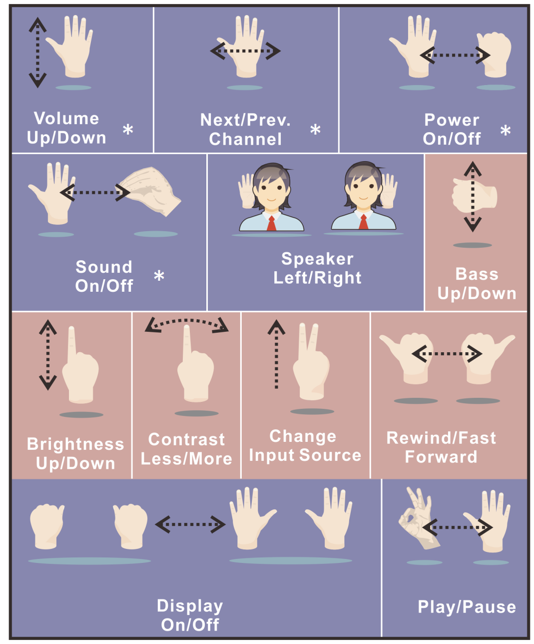

Figure 6. Gesture set with single gestures. * denotes gestures inspired from [37]. Other gestures were newly created by participants. We distinguish between abstract (light orange background) and non-abstract (dark purple background) gestures. Note that the two-handed gesture used for "Display On/Off" is a single gesture-command mapping.

scenario with single gestures increases by three for each new objects (Figure 5, solid line). When combining gestures, the best case scenario involves only the creation of one signifier gesture, with operator gestures being reused across all signifiers (Figure 5, dashed line). The worst case scenario involves the creation of one signifier gesture, and three new operator gestures for each new object (Figure 5, dotted line). In between exist several possibilities depending on how users would define and reuse operator gestures. During the workshop session, participants managed to design a gesture set with 7 gestures, corresponding to the best case scenario by fully reusing gestures across signifiers. They even went a step further and proposed the "Default" signifier as optional, hence leading to only 6 gestures.

Figure 6 shows the gesture set for standard single gesture interaction. When there were no previous gestures defined by past work [37], the gestures were newly defined by the group discussion. We categorized the single gestures into two types: abstract and non-abstract gestures. We based this categorization on the definition provided in previous work [43]: gestures are abstract when they have no clear association with the commands. For instance, it is not clear why users defined fist or some specific finger(s) for "Increase Bass" or "Increase Brightness".

Interestingly, even with a gesture set limited to only 12 gestures, participants had to define 5 out of 12 gestures as abstract gestures. Abstract gestures will likely impose an extra-cognitive load for end-users to learn and recall. We investigate this aspect in our second experiment.

Feedback and design Considerations

Regarding the **combination** aspect, participants generally agreed that combining gestures allows users to reuse gestures and to remember fewer gestures. Thus, participants clearly detected the potential of the main theoretical advantage of combining gestures. To build on this idea, we propose that the design of a gesture set should consider high-level operator description such as "Increase X", X being then determined by the signifier command. This approach could leverage the reuse of operator gestures across signifiers. In addition, participants proposed a way of merging the concept of single gestures and gestures combination by introducing the "Default" command (i.e. no signifier) for the most frequently used commands. Lastly, participants raised the concern of potential fatigue. All participants mentioned that they would sometimes feel too tired (or lazy) to perform two-handed gestures. This concern highlights the trade-off of combining gestures: cognitive (fewer gestures to recall) versus motor efforts (two gestures to execute). Hence, participants recommended leveraging gestures that would be more comfortable and quicker to perform.

Regarding the **gesture** aspect, participants reported that all gestures should be comfortable for the user to perform. For example, participants suggested that designers should avoid using dynamic gestures on both hands at the same time since moving two hands in two different paths could be very difficult. Participants recommended that designers use static gestures for signifier gestures with the non-dominant hand, and use dynamic or static gestures for operator gestures with the dominant hand. This is in accordance with Guiard's kinematic chain theory [12]: the non-dominant hand serves as coarse frame of reference (i.e. define the signifier), while the dominant hand performs detailed work (i.e. define the operator).

Regarding the **temporal relationship** aspect, participants discussed and questioned the necessity of gesturing both hands simultaneously. Although using both hands could lead to faster interaction time than when using only one hand, it was recommended for designers to take into account users who might prioritize comfort over speed, allowing them to perform a sequential combination using only one hand. Participants also reported they would likely execute the operator gestures just slightly after the signifier gestures - hence a *concomitant* combination. Thus, although not clearly stated, participants expected a *relaxed* temporal relationship. They did not consider a strict version and its potential benefits. We hypothesize that this result is an effect of the TV interaction context and the commands used in the study, which did not include commands requiring parameter(s). For instance, participants considered relative commands (e.g., "Increase/Decrease Volume"), but no absolute commands (e.g., "Set Volume to 80%").

Regarding the **spatial relationship** aspect, participants did not raise any suggestion. However, the created gesture set reveals that the "Sound" signifier is linked to the ear location. Thus, all sound-related commands will have a reference frame *partially relative to an external element* (i.e. a ear). This relative reference frame is not meant to transmit an additional parameter, but to enhance the memorability of the gesture by adding a semantic connotation to the gesture location itself.

Discussion

Figure 7 shows our design space modified to encode our findings with Sankey diagrams. A Sankey diagram allows us to represent the links between design options (colored path),

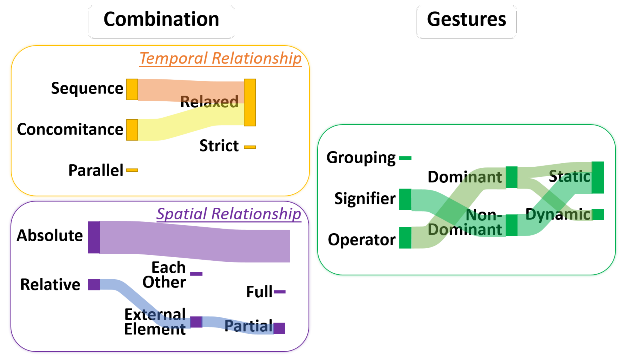

Figure 7. Sankey diagrams visually representing Study 1 findings in our design space.

the proportions in which these links occur (paths width) and the proportions in which options are chosen (option height). Note that such representation does not allow to visualize the specifics of a particular combination, but rather its high-level description via its proportion of option associations.

For the temporal aspect of the combination, the gesture set demonstrates combination of any nature type, i.e. sequential, concomitant, or parallel. However, participants reported a preference for sequential and concomitant temporal relationships, while mentioning the advantages of relaxed temporal constraints. For the spatial aspect of the combination, although participants did not openly mentioned any preferences, the gesture sets reveals that only three combinations out of twelve would use a reference frame partially relative to an external element (an ear). Lastly, participants clearly stated their preferences regarding the gesture properties: a static gesture with the non-dominant hand for the signifiers, and a static or dynamic gesture with the dominant hand for the operators.

Note that these results are restricted to our TV interaction context, and from an initial exploration by end-users of a large design space describing a concept in its infancy stage. We argue that one should refrain from any generalization, and expect properties not explored by this initial study to have valuable potential to explore in future work.

STUDY 2: MEMORABILITY

The design workshop uncovers that 5 out 12 single gestures could be classified as abstract gestures. This suggests that participants were likely to run out of ideas regarding potential intuitive gesture candidates. Study 2 seeks to investigate the memorability of the two gesture sets defined in Study 1, especially regarding the abstract gestures already introduced.

Participants

Fourteen university students and staff (3 females, one left-handed, age 21 to 32, M = 26.8) volunteered for the study. All had knowledge of freehand gestures (e.g., Kinect). Four reported having experience with smart TVs. All participants in Study 2 had not participated in Study 1.

Design

Participants were tasked to memorize two gesture sets - single and gestures combination. Gesture sets presentation was fully counterbalanced across participants. Referents presentation was randomized for each participant.

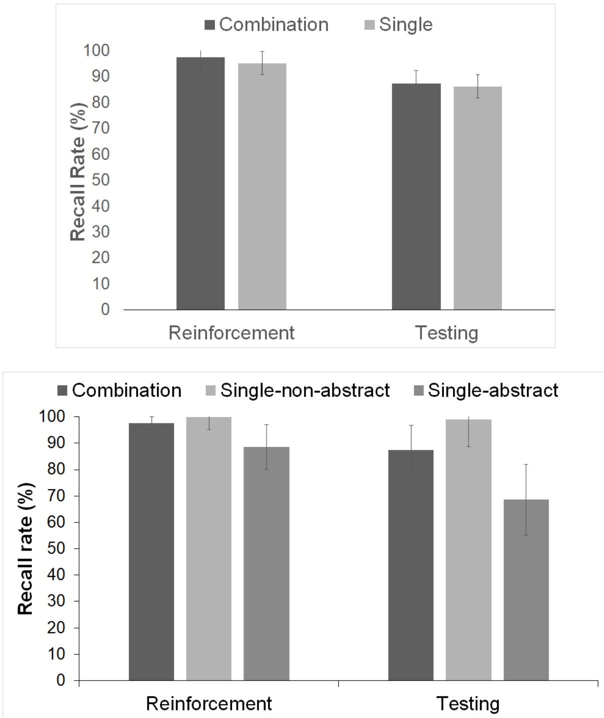

Figure 8. Recall rate during the reinforcement and testing phases (Error bars represent 95% CI).

Participants took part in a three consecutive days experiment following a Learn-Reinforce-Test design [22]. On the first day, we trained participants on one gesture set (Learn). Immediately after that, we asked participants to recall gestures corresponding to the given commands (Reinforce). On the second day, they performed the recall test of the gesture set again (Test). On the same day, participants were asked to learn the second gesture set, and then go through a similar Reinforcement phase. On the third day, memorability was tested for the second gesture set.

Learning Phase

Participants were asked to learn and remember a set of gestures by watching a video explaining the command name, the associated gesture, and its design rationale. Participants could play the video until they had mastered the current gesture. We also provided a crib-sheet (similar to Figures 4 and 6). Finally, to confirm that participants understood the gestures correctly, the experimenter asked them to reproduce the gesture shown in the video and instructed them to watch the video again if they got it wrong. For the gestures combinations, videos demonstrated two-handed relaxed concomitant gestures.

Reinforcement Phase

On the same day, participants were asked to recall the gesture set. The purpose of the reinforcement phase was to test participants' short-term memory and enhance their memory for the next day testing. Participants were required to perform the gestures according to commands randomly presented on the screen. The experimenter would notify participants in case of incorrect number of fingers/hands, trajectories, or poses.

Next-day Testing Phase

The procedure was similar to the Reinforcement phase, except that participants were not told the correctness of their gestures

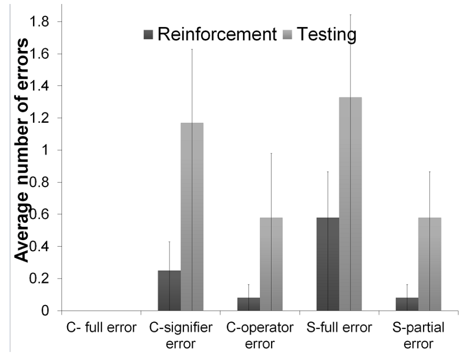

Figure 9. Errors for single (S) and combination (C) of gestures during the reinforcement and testing phases (Error bars represent 95% CI).

to prevent possible cross-recall effects. The experimenter could ask participants to re-perform a gesture for confirmation.

Evaluation Metrics

We considered the number of correctly recalled gestures in terms of poses, paths, and the number of fingers/hands. We further categorized different types of errors [22] to discuss relevant aspects of memorability according to the specifics of the gestures used (single or combination). All measures were collected in both Reinforcement and Testing phases.

For single gestures, we identified two types of relevant errors: mapping errors and partial errors. The former means that participants completely failed to perform the target gesture. The latter is for gestures that were incorrect only in certain parts such as the pose, path, or the number of fingers/hands. This helps us differentiate two types of memorability problems: mapping problems (inability to map the gesture with the command) and partial gesture problems (inability to recall the specifics of the gesture). This will help us determining the effect of abstract gestures in the gesture set.

For gestures combination, since gestures were simple regarding their specifics, we did not distinguish between mapping and partial errors. Instead, we were interested in the combination aspect, and identified three types of recall errors: both hands errors, signifier errors, and operator errors. Both hands errors indicate that participants completely failed to perform the correct gesture with both hands. Signifier and operator errors indicate that participants performed only one correct gesture on one hand but not the other. This will help us determining if users have difficulty with the signifier, the operator, or the combination as a whole.

Results

Our data did not satisfy the normality and the homogeneity of variances assumptions. We hence performed our analysis with Friedman and two-tailed Wilcoxon Sign Rank tests.

Reinforcement Phase

The average recall rates were 95% for the single gesture set and 97% for the gestures combination set (Figure 8), without significant difference. To further compare the user performance of the two gesture sets, we analyzed the recall rates of the single non-abstract and single abstract gestures. The average recall rates were 100% for the single non-abstract gesture set and 88.6% for the abstract ones. This validates our hypothesis: abstract gestures can negatively impact the learnability of a gesture set. We further compared the recall rates of the three gesture types (combination, single abstract and single non-abstract) with a Wilcoxon Signed Ranked Test. The results showed that there was a significant difference between the gestures combinations and the single abstract gestures ($Z = -1.63$, $p < 0.05$), but not between the other pairs. Hence, combining gestures is a viable solution to prevent designing too many single gestures, potentially leading to abstract gestures.

To further understand this results, we analyzed the number of errors per error type (Figure 9). Single gestures lead to more mapping errors ($M = 0.58$, $SD = 0.99$) than partial errors ($M = 0.08$, $SD = 0.29$). A Wilcoxon Signed Rank Test confirmed the significant effect ($Z = -1.86$, $p < 0.05$). Hence, memorability problems, due to abstract gestures, mostly arise from the *mapping* rather than the *specifics* of the gestures.

For gestures combinations, participants did not have any both-hands error. However, there were some signifier errors ($M = 0.25$, $SD = 0.62$) and operator errors ($M = 0.08$, $SD = 0.29$). A Friedman Test shows no significant effect between signifier and operator errors. Thus, we can assume that errors are not linked to the hierarchical construction of the combination.

Next-day Testing Phase

The average recall rates were 86% for the single gestures and 87% for the gestures combinations. A Wilcoxon Signed Rank Test shows no significant effect between the two conditions. Note that both gesture sets suffer from a significant drop of memorability performance between the Reinforcement phase and the Testing phase: from 95% to 86% for single gestures ($Z = 2.41$, $p < 0.05$), and from 97% to 87% for gestures combinations ($Z = 2.01$, $p < 0.05$).

The average recall rate were 98% for the single non-abstract gestures and 69% for the abstract ones. A cross comparison using Wilcoxon Signed Rank test showed significant differences between combinations and single abstract gestures ($Z = -1.66$, $p < 0.05$), and between the two types of single gestures ($Z = -2.04$, $p < 0.05$). Thus, like the Reinforcement phase, abstract single gestures have a negative impact on memorability.

Single gestures have on average more mapping errors ($M = 1.33$, $SD = 1.78$) than partial errors ($M = 0.58$, $SD = 0.99$). A Wilcoxon Signed Rank Test confirmed a significant effect ($Z = -2.07$, $p < 0.05$), also confirming that abstract gestures are problematic because of the gestures/commands mapping.

Gestures combinations have no both-hands error, but lead to signifier errors ($M = 1.17$, $SD = 1.59$) and operator errors ($M = 0.58$, $SD = 1.38$). A Friedman Test confirmed a significant effect ($\chi^2(2) = 6.40$, $p < 0.05$). In contrast to the Reinforcement phase, it appears that signifier gestures lead to slightly less recall than operator gestures. Participants reported that they intuitively associated the "Ear" gesture with "Volume up/down". Instead, "Volume up/down" is a frequently-used command associated with the "Default" signifier.

Discussion

As expected, when categorized as abstract, single gestures lead to memorability problems: participants struggled remembering the *mapping* of abstract gestures. The experiment involved only 5 abstract gestures out of 12. It is likely that as the number of command increases, it will become increasingly difficult for designers to come up with new single gestures without creating abstract gestures. Thus, users may experience even more difficulty memorizing real gesture sets containing several abstract gestures than what our experiment revealed.

Combining gestures lead to improved memorability compared to single abstract gestures. This validates the main rationale of our interest for combining gestures. However, like single gestures, gestures combinations suffer from a drop of performance between the Reinforcement and Testing phases. The Testing phase revealed a significant difference between signifier errors and operator errors. There are two possible explanations for this result. First, it can simply be due to the actual gestures defined for each category (signifier and operator). In this case, the result simply reveal a memorability problem between two gesture sets. Second, this difference can reveal the usage of two mental models: one used for the signifiers, and one used for the operators.

STUDY 3: GESTURE PREFERENCE

Although it can be somewhat premature while gestures combination is a new concept in its early stage, we seek to investigate if end-users can perceive the limitations of single gestures and the potential advantages of combining gestures. Such result can prove that the concept and its newly open research questions are worth investigating.

We use a Wizard-of-Oz approach to simulate TV interaction, so as to not affect users' perception due to gesture recognition failures of a specific algorithm. As with other Wizard-of-Oz studies, participants were not told of the presence of wizards.

Participants

Another fourteen university students and staff (four females, one left-handed, age 20 to 35, M = 26.1) volunteered for the study. They all had knowledge about free-hand gestures, and three reported to have experience with smart TVs.

Apparatus and setup

A SHARP PN-L602B 60-inch large display was placed vertically in front of a sofa. Behind the display, the wizard used a PC (2 GHz Intel Xeon CPU PC with Windows 7) connected to the large display. To simulate TV watching experiences, we compiled a list of 5-10 minutes TV video clips (e.g., music videos, drama, live show) captured prior to this study. We did not use a real TV given that we wanted to control what scenarios users would encounter so as to cover all 12 commands compiled in Study 1. The wizard used keyboard shortcuts, an audio driver software (Realtek High Definition Audio), and a graphic driver software (Intel HD Graphics) to execute the commands according to participants' gestures. Gestures were captured in real time by a webcam (Logitech HP Pro C920) installed on the top of the large display.

Procedure

First, participants were asked to sit on the sofa and then provided with an introductory video of existing smart TV systems (similar to Study 1). After that, the experimenter introduced the two gesture sets (single gestures and gestures combinations) and their design rationales. The gesture sets presentation was fully counterbalanced across participants. We requested participants to first get familiar and memorize the gesture set. The experimenter read out a sequence of scenarios (e.g., "Your friend is

Participant No	Volume up/down	Next/Prev Ch.	Power on/off	Bass up/down	Speaker left/right	Sound on/off	Brightness up/down	Contrast more/less	Display on/off	Change Input Source	Rewind/FFW	Play/Pause
P1*	S	S	S	C	S	C	C	C	C	C	C	C
P2	S	S	S	S	S	C	S	S	S	C	S	S
P3	S	S	S	S	S	S	C	C	C	S	C	S
P4	S	S	S	C	C	S	C	C	C	C	C	S
P5	S	S	S	C	S	C	C	C	C	C	C	S
P6*	S	S	S	C	S	S	C	C	C	C	C	S
P7	S	S	S	C	S	C	C	C	C	C	C	S
P8	S	S	S	S	S	S	S	S	S	S	S	S
P9	S	S	S	S	S	S	C	S	S	C	S	S
P10	S	S	S	S	S	S	C	C	S	S	S	S
P11	S	S	S	C	C	S	S	S	S	S	C	S
P12	S	S	S	C	S	S	C	C	C	C	C	S
P13*	S	S	S	C	S	C	C	C	C	C	C	S
P14	S	S	S	S	S	S	S	S	S	S	S	S

Figure 10. Gesture preference (S: Single gestures, C: Combination) across all commands for each participant. * refers to participants who have experiences with smart TVs.

calling you, please input the command 'Sound off'.") and participants would execute the commands accordingly. When participants performed the correct gesture, the wizard would issue the corresponding TV command. Otherwise, the wizard generated a beep sound to indicate an error. The experimenter informed participants of the correct gesture after finishing the experiment. This whole process was repeated for the other gesture set.

After getting familiar with both gesture sets, participants were asked to perform the gestures again with the TV using a process similar to the previous step. This time, participants were allowed to freely choose whether to use a single or a gestures combination for each command. We video recorded participants' choices for further analysis. At the end of the experiment, we interviewed participants regarding their rationale behind each decision and their general perception towards single gestures and gestures combinations.

Findings

We discuss our findings according to: the comparison between single and gestures combinations, the handedness aspect of gesture interaction, and the gesture sets used in our study.

Direct Comparison

Most participants preferred gestures combinations for commands they considered the single gestures counterpart not meaningful: "*The mapping between brightness and index finger [in the single gesture set] doesn't make sense to me.*" (P6). Following the same rationale, participants preferred single gestures when they judged the mapping gesture/command as 'intuitive'. For instance, participants unanimously commented that they preferred single gestures for the "Default" commands (Figure 10, 3 first columns).

These results indicate that participants' preferences is mainly based on single gestures rather than on the combination counterpart. To validate this assumption, we compare participants' preferences with the memorability data collected in Study 2 (Figure 11). We found that the categorization of a gesture into 'abstract' or 'non-abstract' is significantly correlated to user's preferences (Pearson's $r(10) = -0.83$, $p < 0.001$). This categorization is also correlated to the memorability performance (Pearson's $r(10) = -0.82$, $p < 0.01$). Thus, by transition, preferences are also significantly correlated to the memorability of the gesture set (Pearson's $r(10) = 0.73$, $p < 0.01$). Interestingly, participants' preferences are non-significantly correlated to the memorability aspect of the combination gesture set (Pearson's $r(10) = 0.23$, $p = 0.47$).

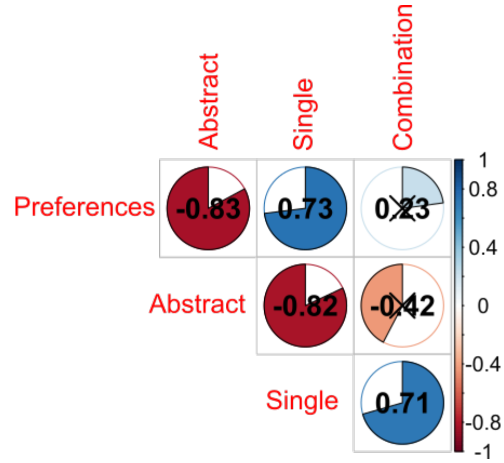

Figure 11. Correlation matrix between users preferences (Study 3) and memorability of single or gestures combinations (Study 2).

Handedness and Gesture Interaction

Some participants reported the desire to use only one hand in order to prioritize comfort over command efficiency (P8, P14). The reasons for using one-handed gestures were many, e.g., eating while watching TV, prioritizing minimized effort, or non-visual command execution (interaction without looking at the TV). For instance, P8 explained that "*I preferred just using one hand so that I can watch TV while eating snacks.*". P14 explained that "*Sometimes I am in a hurry switching between watching TV and other tasks like taking phone, and thus I want to use one hand only so it allows me to quickly manipulate the TV without looking at it.*".

Because of our decision to not reveal the entire design space to participants, gestures combinations were mainly considered as *parallel* temporal combinations. With *sequential* combinations, participants would not have to choose between one-handed and two-handed interactions. During the interviews, P8 and P14 said that if users could perform gestures combinations with one hand sequentially, they may prefer using gestures combinations for some commands.

Gesture Sets

We recognized that the results can vary depending on the gesture design. Thus, to verify the potential influence of the gesture design on the participants' preferences, we asked whether they would design the gestures differently.

14 out of 15 participants would redesign single gestures, particularly commenting on the abstract gestures ("Bass up/down", "Brightness up/down", "Contrast more/less", "Change Input Source" and "Rewind/Fast Forward"). However, only three

participants could come up with an alternative gesture – "Move Hand left/right twice" – for "Rewind/Fast Forward".

For gestures combinations, most participants got "Volume up/down" mixed up with the "Ear" signifier, and "Brightness up/down" with "Contrast more/less". This result is consistent with Study 2. However, no participant came up with any alternative gestures for "Volume up/down", and they suggested that operator gestures for brightness and contrast should support user customization.

Discussion

We have confirmed that combining gestures can effectively complement single gestures. Particularly, we found that a gestures combination is useful when there is no meaningful mapping between a single gesture and a certain command. Our results showed that all participants except two have designated certain combinations for certain commands, indicating the usefulness of combining gestures. However, these two participants stated that they would revise their preferences once we revealed other temporal combination options.

In sum, participants' rationale on whether to use gestures combinations is primarily based on two factors: meaningful mapping of the single gestures, and user comfort.

GENERAL DISCUSSION

We have found that combining gestures allows for meaningful gestures/commands mappings, and can also significantly reduce the size of the gesture set. We also found that users can correctly associate and remember signifiers and operators gestures. Lastly, users tend to prefer gestures combinations over single gestures when single gestures reach their abstract gesture/command mapping.

Although gestures combination was generally well-received, our studies identified a key limitation: combining gestures leads to extra physical effort. Thus, it was recommended by participants to mix both single and gestures combination for practical purposes. However, we contrasted single and gestures combination for comparative purpose only. The goal was not to completely replace the use of single gestures, but to compare several aspects of a novel concept to a known baseline, and to identify future use-cases in which combining gestures will be useful.

LIMITATIONS AND FUTURE WORK

We proposed illustrative examples for each design options in our design space (e.g., transmit a parameter along with the command). These examples can be extended, and then empirically evaluated. We did not choose this option for this initial exploration since (1) we wanted to introduce and to know if the concept was worth exploring, and (2) we could not compare these new input scenarios to single gestures as there is no standard solution to use single gestures with parameters. We believe our work validates further exploration of our design space generative power: We are now planning to create novel ways of inputing commands and/or parameters to demonstrate the full interaction potential of combining gestures.

Regarding interaction scenarios, we used the TV interaction context as a case study. A straightforward extension of this work resides in the use of other interaction scenarios. This will be useful for two reasons. First, results can differ according to user's activity (e.g., watching TV on a sofa or cooking diner). Second, other scenarios can include more gestures in the gestures combinations. An interesting research area is the common breadth/depth tradeoff of a hierarchy structure [21]. For now, we restricted our initial exploration to combinations of only single gestures (depth=1) and two gestures (depth=2) for 12 commands. We are interested in investigating the evolution of the memorability and preference metrics with hierarchies with different breadth and depth.

Regarding the training procedure, prior to Study 2, we performed several pilot studies trying to ask users to memorize gestures combinations using different training procedure. We found that teaching users each gesture one by one was not effective. This is because participants could not understand the whole gesture-command mapping. We found that providing a crib-sheet similar to Figure 4 can greatly assist users to memorize gestures. The literature proposes several more efficient gesture guiding systems to help novices execute and learn single-handed gestures [9] or symmetric two-handed gestures [27] (i.e. one command triggered with both hands). A gestures combination brings new challenges for a gesture guiding system depending on the different *temporal* and *spatial* combinations offered by the design space.

CONCLUSION

Mid-air hand gestures can be used to control augmented home appliances in smart environments. Previous work mainly focuses on one-to-one gesture-command mapping. However, as the number of objects and their interactive functionalities increase, it becomes difficult to define new gestures without including memorability issues: mapping and gesture specifics. The goal of this research was to identify and provide a way to extend single one-handed gestures to gestures combinations. This complementary approach accommodates the reuse of gestures by creating a hierarchy of commands to trigger.

We first propose a design space for describing the temporal and spatial aspects of the combination, and also the gestures type involved in the combination. The generative power of the design space has the potential to lead to new ways of inputing commands and their parameter(s) via gesture interaction.

We then validate our experimental setup via pilot studies. We show that the TV interaction context is suitable to study gesture interaction. In addition, a design workshop is preferable to a standard elicitation study considering the novelty and the large number of design options of combining gestures.

We next report the results from three studies to investigate how end-users would design gestures combinations, the memorability of a gesture set composed of combination, and users' preferences. Results show that end-users mostly focus on *temporal* aspects of the combination and manage to optimize the reuse of gestures. In addition, gestures combinations improve the memorability of a gesture set compared to abstract single gestures, i.e. gestures without a clear mapping with their associated command. Lastly, users' preferences go toward gestures combinations when the single gesture counterparts are abstract.

ACKNOWLEDGMENT

We acknowledge the support from the JSPS fellowship program, and thanks Dr. NIKSIRAT for his drawings.

REFERENCES

[1] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindlbauer, Alexandra Ion, Shengdong Zhao, and Jeffrey Tzu Kwan Valino Koh. 2012. *Understanding Mid-Air Hand Gestures : A Study of Human Preferences in Usage of Gesture Types for HCI*. Technical Report. 10 pages.

[2] Ravin Balakrishnan and Ken Hinckley. 2000. Symmetric Bimanual Interaction. In *Proc. CHI 2000*. ACM, 33–40. DOI:http://dx.doi.org/10.1145/332040.332404

[3] Ravin Balakrishnan and Gordon Kurtenbach. 1999. Exploring Bimanual Camera Control and Object Manipulation in 3D Graphics Interfaces. In *Proc. CHI 1999*. ACM, 56–62. DOI: http://dx.doi.org/10.1145/302979.302991

[4] Amartya Banerjee, Jesse Burstyn, Audrey Girouard, and Roel Vertegaal. 2011. Pointable: An In-air Pointing Technique to Manipulate Out-of-reach Targets on Tabletops. In *Proc. ITS 2011*. ACM, 11–20. DOI: http://dx.doi.org/10.1145/2076354.2076357

[5] William Buxton and Brad Myers. 1986. A Study in Two-handed Input. In *Proc. CHI 1986*. ACM, 321–326. DOI:http://dx.doi.org/10.1145/22627.22390

[6] Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. 1991. A morphological analysis of the design space of input devices. (apr 1991). DOI: http://dx.doi.org/10.1145/123078.128726

[7] Dipen Dave, Ashirwad Chowriappa, and Thenkurussi Kesavadas. 2013. Gesture Interface for 3D CAD Modeling using Kinect. *Computer-Aided Design and Applications* 10, 4 (2013), 663–669. DOI: http://dx.doi.org/10.3722/cadaps.2013.663-669

[8] William Delamare, Céline Coutrix, and Laurence Nigay. 2015. Designing guiding systems for gesture-based interaction. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems - EICS '15*. ACM Press, Duisburg, Germany, 44–53. DOI:http://dx.doi.org/10.1145/2774225.2774847

[9] William Delamare, Thomas Janssoone, Céline Coutrix, and Laurence Nigay. 2016. Designing 3D Gesture Guidance : Visual Feedback and Feedforward Design Options. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*. ACM Press, New York, New York, USA, 152–159. DOI: http://dx.doi.org/10.1145/2909132.2909260

[10] Simon Fothergill, Helena Mentis, Pushmeet Kohli, and Sebastian Nowozin. 2012. Instructing people for training gestural interactive systems. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, New York, New York, USA, 1737–1746. DOI: http://dx.doi.org/10.1145/2207676.2208303

[11] William Freeman and Craig Weissman. 1995. Television control by hand gestures. In *Proc. IEE Intl. Wkshp. on Automatic Face and Gesture Recognition*.

[12] Yves Guiard. 1987. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior* 19, 4 (1987), 486–517.

[13] Michael Hoffman, Paul Varcholik, and Joseph J. LaViola. 2010. Breaking the status quo: Improving 3D gesture recognition with spatially convenient input devices. In *2010 IEEE Virtual Reality Conference (VR)*. IEEE, 59–66. DOI:http://dx.doi.org/10.1109/VR.2010.5444813

[14] Maria Karam and m. c. Schraefel. 2005. A Taxonomy of Gestures in Human Computer Interactions. (2005), 1–45. DOI:http://dx.doi.org/10.1.1.97.5474

[15] Juha Kela, Panu Korpipää, Jani Mäntyjärvi, Sanna Kallio, Giuseppe Savino, Luca Jozzo, and Sergio Di Marca. 2006. Accelerometer-based gesture control for a design environment. *Personal and Ubiquitous Computing* 10, 5 (aug 2006), 285–299. DOI: http://dx.doi.org/10.1007/s00779-005-0033-8

[16] Elsa Kirchner, JosÃľ FernÃąndez, Peter Kampmann, Martin SchrÃűer, Jan Metzen, and Frank Kirchner. 2015. *Intuitive Interaction with Robots–Technical Approaches and Challenges*. 224–248. DOI: http://dx.doi.org/10.1007/978-3-658-09994-7_8

[17] Christine Kühnel, Tilo Westermann, Fabian Hemmert, Sven Kratz, Alexander Müller, and Sebastian Möller. 2011. Im home: Defining and evaluating a gesture set for smart-home control. *International Journal of Human Computer Studies* 69, 11 (oct 2011), 693–704. DOI: http://dx.doi.org/10.1016/j.ijhcs.2011.04.005

[18] G. Kurtenbach, T. P. Moran, and W. Buxton. 1994. Contextual Animation of Gestural Commands. *Computer Graphics Forum* 13, 5 (dec 1994), 305–314. DOI: http://dx.doi.org/10.1111/1467-8659.1350305

[19] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-air Interaction. In *Proc. UIST 2015*. ACM, 63–71. DOI: http://dx.doi.org/10.1145/2807442.2807489

[20] Frieder Loch. 2012. Hierarchical Gestures: Gestural Shortcuts for Touchscreen Devices. (2012), 1–124. http://friederloch.de/2011/thesis/Thesis/thesis.pdf

[21] Dwight P. Miller. 1981. The Depth/Breadth Tradeoff in Hierarchical Computer Menus. *Proceedings of the Human Factors Society Annual Meeting* 25, 1 (oct 1981), 296–300. DOI: http://dx.doi.org/10.1177/107118138102500179

[22] Miguel A Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-designed and User-defined Gesture Sets. In *Proc. CHI 2013*. ACM, 1099–1108. DOI: http://dx.doi.org/10.1145/2470654.2466142

[23] Mathieu Nancel, Julie Wagner, and Emmanuel Pietriga. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proc. CHI 2011*. ACM, 177–186. DOI: http://dx.doi.org/10.1145/1978942.1978969

[24] Vijayakumar Nanjappan, Hai-Ning Liang, Feiyu Lu, Konstantinos Papangelis, Yong Yue, and Ka Lok Man. 2018. User-elicited dual-hand interactions for manipulating 3D objects in virtual reality environments. *Human-centric Computing and Information Sciences* 8, 1 (dec 2018), 31. DOI: http://dx.doi.org/10.1186/s13673-018-0154-5

[25] François Portet, Michel Vacher, Caroline Golanski, Camille Roux, and Brigitte Meillon. 2013. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Personal and Ubiquitous Computing* 17, 1 (2013), 127–144.

[26] Asm Mahfujur Rahman, M. Anwar Hossain, Jorge Parra, and Abdulmotaleb El Saddik. 2009. Motion-path based gesture interaction with smart home services. *Proceedings of the seventeen ACM international conference on Multimedia - MM '09* (2009), 761–764. DOI:http://dx.doi.org/10.1145/1631272.1631407

[27] Gustavo Rovelo, Donald Degraen, Davy Vanacken, Kris Luyten, and Karin Coninx. 2015. Gestu-Wan - An Intelligible Mid-Air Gesture Guidance System for Walk-up-and-Use Displays. In *Proc. Interact '15*. 368–386. DOI: http://dx.doi.org/10.1007/978-3-319-22668-2_28

[28] Quentin Roy, Sylvain Malacria, Yves Guiard, Eric Lecolinet, and James Eagan. 2013. Augmented Letters: Mnemonic Gesture-based Shortcuts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2325–2328. DOI: http://dx.doi.org/10.1145/2470654.2481321

[29] Marcos Serrano, Eric Lecolinet, and Yves Guiard. 2013. Bezel-Tap Gestures: Quick Activation of Commands from Sleep Mode on Tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3027–3036. DOI: http://dx.doi.org/10.1145/2470654.2481421

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

[31] Radu Daniel Vatavu. 2012a. Nomadic gestures: A technique for reusing gesture commands for frequent ambient interactions. *Journal of Ambient Intelligence and Smart Environments* 4, 2 (2012), 79–93. DOI: http://dx.doi.org/10.3233/AIS-2012-0137

[32] Radu-Daniel Vatavu. 2012b. User-defined gestures for free-hand TV control. In *Proc. EuroiTV 2012*. ACM, 45–48. DOI:http://dx.doi.org/10.1145/2325616.2325626

[33] R.-D. Vatavu. 2013a. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *Journal of Ambient Intelligence and Smart Environments* 5, 2 (2013), 187–211.

[34] Radu-Daniel Vatavu. 2013b. There's a world outside your TV: exploring interactions beyond the physical TV screen. In *Proc. EuroiTV 2013*. ACM, 143–152. DOI: http://dx.doi.org/10.1145/2465958.2465972

[35] Radu-Daniel Vatavu. 2019. The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19* (2019), 1–13. DOI:http://dx.doi.org/10.1145/3290605.3300454

[36] Radu-Daniel Vatavu and Stefan-Gheorghe Pentiuc. 2008. Interactive Coffee Tables: Interfacing TV within an Intuitive, Fun and Shared Experience. In *Proc. EuroiTV 2008*. Springer Berlin Heidelberg, 183–187. DOI: http://dx.doi.org/10.1007/978-3-540-69478-6_24

[37] Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. 2014. Leap gestures for TV: insights from an elicitation study. In *Proc. TVX 2014*. ACM, 131–138. DOI: http://dx.doi.org/10.1145/2602299.2602316

[38] RD Radu-Daniel Vatavu, CM Catalin-Marian Chera, WT Wei-Tek Tsai, and CM Catalin-Marian Chera. 2012. Gesture profile for web services: an event-driven architecture to support gestural interfaces for smart environments. *Ambient Intelligence* (2012), 161–176. DOI:http://dx.doi.org/10.1007/978-3-642-34898-3_11

[39] Frederic Vernier and Laurence Nigay. 2001. A framework for the combination and characterization of output modalities. *Interactive Systems Design, Specification, and Verification* (2001), 35–50. DOI: http://dx.doi.org/10.1.1.32.9133

[40] Vogiatzidakis and Koutsabasis. 2019. Frame-Based Elicitation of Mid-Air Gestures for a Smart Home Device Ecosystem. *Informatics* 6, 2 (2019), 23. DOI: http://dx.doi.org/10.3390/informatics6020023

[41] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, New York, New York, USA, 841–850. DOI: http://dx.doi.org/10.1145/2470654.2470774

[42] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009a. User-defined Gestures for Surface Computing. In *Proc. CHI 2009*. ACM, 1083–1092. DOI: http://dx.doi.org/10.1145/1518701.1518866

[43] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009b. User-defined Gestures for Surface Computing. In *Proc. CHI 2009*. ACM, 1083–1092. DOI: http://dx.doi.org/10.1145/1518701.1518866