# User-defined gesture set for home automation control

Charbel Francisco Elias (Charbel-Francisco.Elias@grenoble-inp.org), Milena Marković (Milena.Markovic@grenoble-inp.org)

January 2021

## 1. Introduction

Home automation systems are becoming an increasingly common and widely available form of technology. With this expansion of availability comes a diverse new body of potential users. Currently, home automation systems are most commonly controlled through digital applications, remotes or voice controlled interfaces. These options come with their advantages and disadvantages: knowledge of how to use digital applications and remotes can be hard to acquire for some demographics and voice controlled interfaces come with concerns of privacy as well as limitations to the available languages. Our proposition is to find a set of intuitive hand gestures that can be used for the purpose of home automation control that are easy to learn and execute. The goal of this paper is to conduct a user study in order to propose an input vocabulary for some of the most common tasks in home automation.

## 2. Related Work

Relevant prior work includes studies of gesture evaluation methods, gesture set definition, and user input elicitation.

Correia et. al. [1] analyzes the existing evaluation methods of these gestures based on different aspects such as accessibility, usability, personalization, privacy, ambiguity, or anthropology. A new framework is introduced, which permits the discussion of such aspects in a socio-technical way, within a triadic relationship consisting of 3 dimensions: people, domotics, and gestural interaction. Considering these aspects increases the intuitiveness of smart home automation applications and their controls as the current development of these applications is based on technological aspects and not what is desirable and intuitive for users.

Kuehnel et. al. [5] worked on defining and evaluating a gesture set for smart-home control by first finding the gesture vocabulary then evaluating the gesture set based on the ability to map the gesture to a specific task. Finally, the memorability of the gesture set was evaluated. However, during the user research, a smartphone was used as an input device for the gesture registration which restricts the gesture vocabulary, especially when compared to gesture recognition using computer vision, data gloves, Kinect, EMG devices, or other wearable devices.

# 3. Developing a gesture set

## 3.1 Overview

Following a procedure similar to a study by Wobbrock et al. [3] where they performed user behaviour elicitation in order to develop a set of gestures for surface computing, we conducted a series of interviews to acquire data on what gestures users would intuitively associate with certain actions.

First off, we needed to define the set of referent home automation actions for which we would elicit gestures. First we compiled a list of common smart home devices. This list includes: lights, blinds or curtains, volume and media control, air conditioning and heating, kitchen appliances, smart electrical outlets, security systems and doorbells. We recognized some situations where universal gesture controls wouldn't be appropriate or practical. One of the situations are appliances that are non-standard, too specific, customizable or require complex input. Most smart kitchen appliances fall in this category: smart coffee makers and slow cookers require complex input, smart refrigerators don't have a standard set of functionalities etc. The second situation are actions that require any kind of feedback from the system to the user. This includes security systems and, as their subset, smart doorbells, for which the user requires a screen or speakers to receive images or sound from the appliance and as this adds an additional layer of complexity (ie. how to determine where to deliver this feedback), these actions were also rejected.

Taking this into consideration, we ended up with a set of 7 referent actions: (1) Turning lights on and off, (2) Controlling music playback (play and stop), (3) Playing the next song, (4) Controlling media volume (increase and decrease), (5) Turning air conditioning on and off, (6) Adjusting the air conditioner thermostat, and (7) Lowering and raising window blinds. Two other good candidates were also controlling heating and adjusting positions of curtains, but due to their similarity to air conditioning and window blinds, they were left out. This paper will only focus on matching gestures conceptually to actions to create an input vocabulary without much attention to the implementation of the system, detection of gestures, integration with systems etc. and any constraints imposed by these factors.

## 3.2 Participants

Thirteen participants were recruited for the study. The median age of participants was 25 with ten participants being in their twenties, three in fifties and one in their teens. Five used some form of home automation previously and one used a form of gesture based controls (music control in some BMW cars). None of the participants have any contact with user interface design, three participants work related to computer science (software engineers and one informatics teacher) while the rest include three teachers, a surgeon, a student, engineers from several fields (mechanical, energetics, biomedical, civil) and a public administration manager.

## 3.3 Procedure

The user interviews were conducted through Zoom. Participants were presented with seven home control actions: an oral description and a video recording representing the action, to make sure all participants have the same visual of the action. After presenting each action, the

participant is asked to propose a gesture that they think corresponds to it. After the participant suggests gestures for all actions, they are asked three additional questions about each gesture: What is their reasoning behind suggesting that gesture and any additional comments about it, how they would rate that the gesture matches the action on a scale from 1 to 5 where 1 is a very bad match and 5 is a very good match, and how they would rate the difficulty of performing this action on a scale from 1 to 5 where 1 is very difficult and 5 is very easy.

## 3.4 Results

### 3.4.1 Taxonomic classification of gestures

#### Definition

We will be following the terminology proposed by McNeill [4] that classifies gestures in four classes: iconics, metaphorics, deictics and beats. Iconic gestures bear close formal relationship to a referent (an action, event, phenomenon), and some are used so often they became canonical for depicting certain things. Metaphoric gestures are similar to iconics, but they present abstract concepts though metaphors. Deictic gestures are pointing movements toward objects or abstract spaces to draw attention to them. Beats are movements without any apparent meaning, small movements without specific gesture space.

#### Taxonomic breakdown in our data set

The thing that is first noticeable is that our dataset doesn't contain any deictic gestures. Due to the commands being of imperative nature, this is somewhat expected. In total, most gestures proposed were metaphoric (44,6%), followed by iconic (30,4%) and lastly beat gestures (25%) (figure 1).

The actions that belong to the on/off switch type have the most beat gestures proposed - clapping, snapping fingers etc. This can be explained by these actions being associated with simple real life actions like flipping a switch, pressing a button, or more specifically giving a signal to adjust a binary setting (on/off). For this type of action, beat gestures are followed by metaphorics, gestures that imitate abstract concepts. An example of this is a gesture for turning the light on and off by opening and closing a fist - opening the fist to symbolize turning the light bulb on and radiating light and inversely closing an open palm into a fist to symbolize the diminishing of light when turning it off.

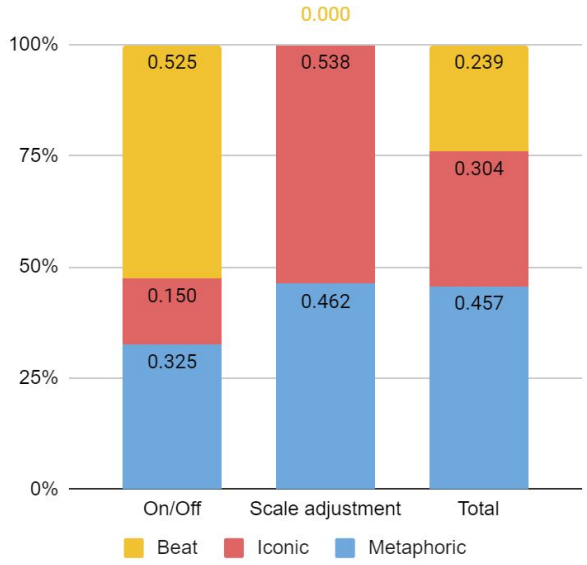


Figure 1. Percentages of gestures in each category for the two different types, and for all gestures.

On the other hand for scale adjustment actions, there are no beat gestures and the user suggestions are almost evenly split between iconic (53,8%) and metaphoric (46,2%) gestures. To illustrate on the example of volume control, an iconic gesture would be moving a finger in a circular motion or pretending to twist a physical "knob" like ones that we would encounter on a radio device, and a metaphoric one would be moving a hand up, symbolizing the increasement of volume and moving the hand down to decrease. Additionally, on figure 2 we can observe a difference in the ratio of iconic and metaphoric gestures between volume control and thermostat, and window blinds.

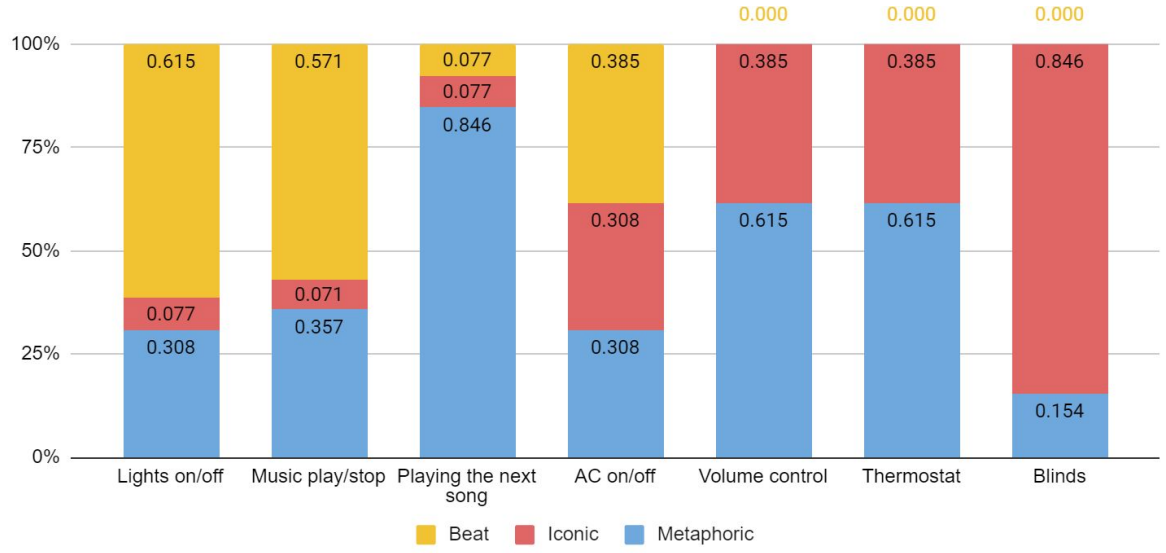


Figure 2. Percentages of gestures in each category for each action.

Since volume control and thermostat adjustments concern adjusting something that is more abstract, more participants proposed metaphoric gestures. On the other hand, adjusting window blinds deals with the actual position of a physical object, so iconic gestures are more prevalent - mostly imitating the motion of physically grabbing and moving the blinds up and down.

### 3.4.2 Gesture agreement

This gesture set is based on the participants' gestures and their reasoning behind each of the performed gestures. After all participants provided their gestures corresponding to each of the 7 actions, we manually grouped the similar gestures taking into consideration both similarity of execution of a gesture, and the reasoning and additional comments about the gesture that the participants provided. For example, if two participants performed a hand swiping motion for the same action with a slight difference in the movement or orientation of the arm, but had the same reasoning behind performing that gesture, we considered these two gestures to be the same.

In order to better evaluate the results, we calculated the agreement rate as introduced by Wobbrock et al. [2] and further utilized in [3]. The agreement rate is a score that reflects the degree of agreement among participants for experimental designs that elicit multiple proposals for the same referent (i.e., the percent of participants suggesting the most popular proposal for a given referent). The agreement rate would increase if more participants perform similar gestures for a single action, and vice versa. In order to calculate the agreement rate for a single action, the total number of participants and the number of participants for each performed gesture is taken into consideration through the following equation (1).

$$A = \frac{\sum_{r \in R} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|}\right)^2}{|R|} \qquad (1)$$

In equation (1), as proposed in [3], r is a referent in the set of all referents R, $P_r$ is the set of proposed gestures for referent r, and $P_i$ is a subset of identical gestures from $P_r$. As an example, consider the agreement for the action of turning the lights on and off. The proposed gesture set included 5 different gestures distributed as follows: 7 participants proposed clapping, 3 proposed closing the hand palm to make a fist, and the 3 remaining gestures were only proposed once. Using this information, as an example we can calculate the agreement rate of Lights on/off (2).

$$A_{Lights\ on/off} = \left(\frac{7}{13}\right)^2 + \left(\frac{3}{13}\right)^2 + \left(\frac{1}{13}\right)^2 + \left(\frac{1}{13}\right)^2 + \left(\frac{1}{13}\right)^2 = 0.36 \qquad (2)$$

The agreement rates for all of the referent actions are shown on chart on figure 3.

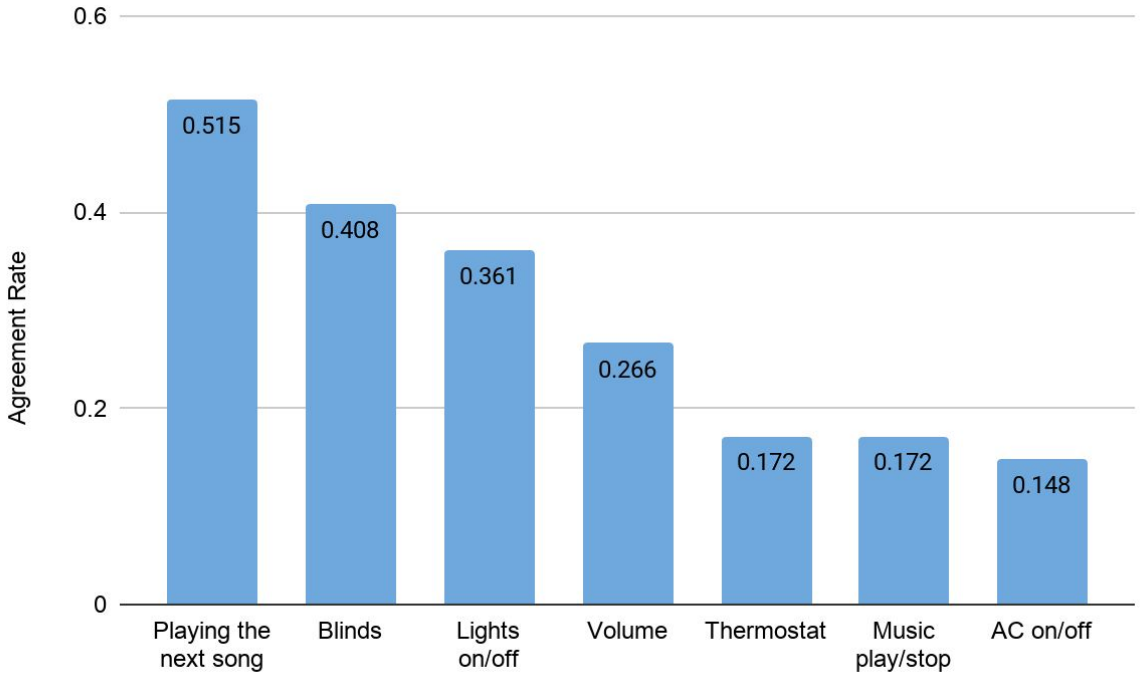


Figure 3. Agreement rate for each action in descending order

The mean agreement of all actions is $A_{mean}$ = 0.28 with a maximum of 0.515 for the referent action *playing the next song*. The gesture set corresponding to this action was the most consistent where 70% of participants suggested swiping left or right with 1 hand. The rest of the suggested gestures were also similar, ranging from finger swiping to finger snapping. The reason behind the consistency of these gestures is that 85% are common metaphoric gestures similar to swiping on a touchscreen phone.

The agreement for *lowering and raising window blinds* and *turning lights on and off* is also considered good with a score of 0.408 and 0.361, respectively. The agreement rate for volume adjustment is lower than the previous three actions due to the fact that the gestures were separated into 2 major sets: 38% of the participants proposed a circular motion with one finger imitating the physical rotation of a knob where rotating clockwise increases the volume and rotating counter-clockwise decreases it, and 30% proposed a hand swipe up and down.

The last three actions are *adjusting air conditioner thermostat, music playback control (play and stop),* and *turning AC on and off*, and their agreement rates are all lower than 0.2 due to the fact that amongst the 13 proposed gestures for a single action, we can find up to 8 different gestures. For the air conditioner related actions, this is likely because they are less common, and because some participants tried to propose gestures that are specific for that physical device while others re-used gestures that they already used for other actions. As an example, for the air conditioner control, we notice on figure 2, the almost equal distribution of the taxonomy of the gestures over the 3 categories: beat, iconic and metaphoric. One of the gestures was to put hands forward to imitate the air flow; another gesture was to move the

finger up and down to imitate the physical swinging of the AC flaps; an interesting gesture was flapping the hand next to the face to generate wind. Other gestures were similar to different ON/OFF actions such as clapping or snapping the finger to turn on or off - in conclusion, a diverse set of gestures which resulted in low agreement rates.

### 3.4.3 Resulting input vocabulary

Based on the agreement rates, the selection of gestures for the input vocabulary is given in table on figure 4.

| Action | Gesture | Agreement rate | Average goodness rating (scale 1-5, 1 bad match, 5 good match) | Average ease of execution rating (scale 1-5, 1 difficult, 5 easy) |
|---|---|---|---|---|
| Turning lights on and off | Clapping | 0.361 | 4.75 | 5 |
| Music playback control (play and stop) | Snapping fingers | 0.172 | 4 | 4.9 |
| Playing the next song | Swiping arm to the right | 0.515 | 4.3 | 4.9 |
| Controlling media volume (increase and decrease) | Circular motion with one finger, clockwise to increase/counter clockwise to decrease | 0.266 | 4.6 | 4.4 |
| Turning AC on and off | Waving hand near face | 0.148 | 4.3 | 4.6 |
| Adjusting air conditioner thermostat | Hand vertical, swiping up and down | 0.172 | 3.6 | 4 |
| Lowering and raising window blinds | Palm horizontal, moving up and down | 0.408 | 4 | 4.75 |

Figure 4. Referent actions with the gestures that were suggested the most and the agreement rate color coded by how high the agreement is (Green - high, yellow - medium, red - low)

We will now discuss each of these matches in order of their agreement rate from highest to lowest.

The one with the highest agreement rate is for action *Playing the next song* and the gesture is *Swiping arm to the right* (figure 5). Out of 13 participants, 9 proposed this gesture or a variant (eg. swiping with just the palm or with the whole arm, palm facing outward or inward). The average subjective rating of how well the gesture matches the action is 4.3/5 and ease of execution 4.9/5. As for the reasoning the users provided for suggesting this

gesture, they were mostly similar: it reminds them of the gesture used on touchscreen phones to “go forward” in various types of applications (eg. swiping through an image gallery, some music playing applications), moving book pages (although this is in the opposite direction). Some explained they imitate the standard symbol for the next song, that is an arrow, or gave more general explanations that in real life this gesture means “move along” and “continue”.

Figure 5. Participant demonstrating *Swiping arm right* gesture.

The next action that had a high agreement rate is for action *Controlling window blinds* and the gesture most often suggested was *Palm horizontal, moving up and down* (figure 6), suggested by 8 out of 13 participants in two variants - with one hand (5) and with both (2). All of the participants had the same association to the real motion of grabbing the blinds and pulling them down or pushing them up. The average subjective rating of how good this match is is lower than for the others at 4/5, and for the ease of execution 4.75/5. We attribute this slight drop in the subjective rating to the fact that most participants were unsure about which gesture to propose due to the action itself being something they don’t do often (automated or manually) or don’t perceive it as something that is automated at all.

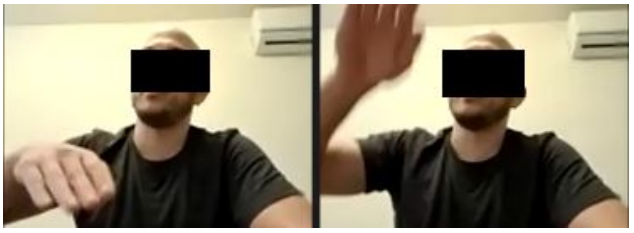

Figure 6. Participant demonstrating *Palm horizontal, moving up and down* gesture.

A similar agreement rate was obtained for the referent action of *Turning lights on and off* and the gesture with most suggestions is *Clapping.* Out of 13 participants, 7 proposed this gesture in two variants - clapping once (5 participants) or twice (2 participants). The average subjective rating of how well the gesture matches the action is 4.75/5 and ease of execution 5/5. All of the participants provided the same reasoning behind this gesture are Clapper systems for light switches that were popular in the 1980s and 1990s and mostly known to the participants by seeing them frequently in films and television shows.

The next action has a lower agreement rate and that is *Controlling media volume (increase and decrease)*. For this action, the best found gesture is *Circular motion with one finger, clockwise to increase/counter clockwise to decrease*. It was suggested by 5 out of 13 participants with the average subjective rating of how well the gesture matches the action being 4.6/5 and ease of execution rating 4.4/5. For most participants, this gesture is an imitation of twisting a physical knob on a radio device with an exception of one participant that has professional musical training and that explained that this is reminiscent of gestures by orchestra conductors.

The last three actions had lowest agreement ratings, so the confidence for whether they are right gestures to be included in this vocabulary is low. Nevertheless, we will discuss them in more detail, starting with the action of *Music playback control (play and stop)*. The most commonly proposed gesture for this action is *Snapping fingers*, by 4 participants out of 13 all of whom provided the reasoning related to finger snapping being commonly used to give or imitate the tempo of music. The subjective rating of how good this match is is 4/5 and ease of execution 4.25/5.

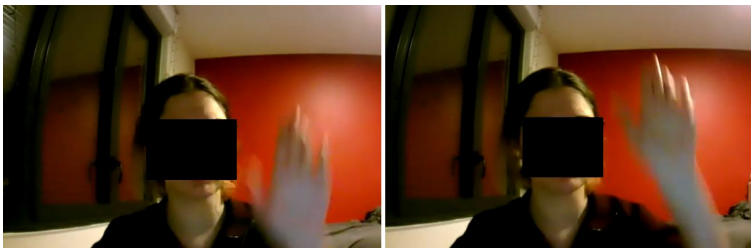

Figure 7. Participant demonstrating *Hand vertical, swiping up and down* gesture.

The second action is *Adjusting the air conditioner thermostat* with the proposed gesture *Hand vertical, swiping up and down* (figure 7) by 3 out of 13 participants. The subjective rating of how good the match is for the action is 3.6/5 and 4.6/5 for the ease of execution. The participants didn't have any conclusive reasoning as to why they performed this gesture except that it's just a gesture for raising and lowering something in everyday life, which can apply here to temperature. This is the reason that it received such a low rating for the subjective goodness: it is ambiguous, not specific to the temperature or air conditioner.
The last action, with the lowest agreement rates was *Turning AC on and off* with the proposed gesture *Waving hand near face*. It was proposed by 4 out of 13 participants in two variants: the first one is the gesture of waving *at* the person's face (fanning), like one is trying to cool themself with the explanation that they are signaling to the air conditioning unit that it's hot and you need to cool down. The second is just waving near the face/head but more with the meaning of waving at the AC itself (which is usually somewhere upwards) as an imitation of what the device does: 1) generates cold air/wind or 2) opens and moves the "flaps". To reiterate, half of the participants that proposed this had in mind to signal what they are feeling so that the automated home responds to accommodate their needs, while the other half meant to imitate what the air conditioner should do (blow wind of cold air or move the flaps of the device).

## 3.4.4 Additional observations

### Number of hands

At the beginning of the interview, participants were told that there are no limitations regarding whether they can use just one or both hands so this was completely up to them. In general, the majority of gestures were one handed, more precisely 78.1% of all suggested gestures which can indicate that users prefer single-handed gestures. The only two common two handed gestures were clapping and two-handed up-down movements with horizontal palms for blind movements. Interestingly, the rest of the two handed gestures were all

suggested by just one single participant that proposed them for 5 out of 7 actions and seemed to favor more complex and non-standard gestures.

### Reversible and toggle gestures

With the exception of the referent action *Play next song**, all actions were reversible, ie. turning something on and off, increasing and decreasing levels of something and moving something in two directions. This means that each action actually requires two gestures - on/increase and off/decrease, and it might be intuitive to use reversible gestures, ie. those that when performed in an opposite direction symbolize the opposite result (eg. rotating left and right, moving up and down, moving left and right, opening and closing).

For the group of on/off gestures (lights and AC) that can also be called toggle gestures, 16/26 proposed gestures were treating this as a "toggle" action, meaning that there is only one single gesture and that its meaning is "change to the opposite state" eg. clapping when lights are turned on turns them off, and vice versa. Of the 10/26 that have different movements that mean on and off, 7/10 were reversible, eg. opening a fist to turn on lights, closing to turn them off.

The referent action for playing and stopping music had the most mismatched gestures: 5/13 had completely different (non-reversible) movements for playing and stopping.

Lastly, for the scale adjustment type of gesture (temperature, volume, window blinds position) there were 34/39 proposed gestures that are a reversible gesture which is, due to the nature of the action expected. One of the interesting examples of proposed gestures for this type that aren't reversible is one provided for adjusting the air conditioning thermostat: the participant proposed that a gesture for decreasing the temperature should be fanning their face (short small wave) with their hand next to their face to signalize they are feeling warm, and for the opposite effect of increasing the temperature, they suggested crossing their arms on their chest to signalize "they are putting on a jacket because it's cold".

Finally, with only 13/78 propositions being different non-reversible gestures for two opposite effects we can conclude that participants preferred and found reversible gestures more preferable, and therefore likely more intuitive.

*(Note: *Play next song* is reversible with an action of playing the previous song, however the participants were specifically asked just about playing the next song so it is excluded from this observation).

## Conclusion

In this paper, we have presented a new set of intuitive gestures to control home automation systems. We performed a user behavior elicitation by conducting a series of interviews with 13 participants with no previous contact with user interface design.

We wanted to conduct these interviews in person, in an environment where the devices were at their disposal, and the participants could see the results of their gestures. Due to the COVID-19 and the restriction of not being able to interview people in person, we were forced to interview the participants through video calls, which may have affected the quality of the interview results.

The participants were presented with oral descriptions of 7 home automation tasks with their relevant videos demonstrating these tasks; then, they proposed the most spontaneous gestures. Starting with the proposed gestures, we grouped the similar gestures based on several factors. Then, we classified the gestures into 3 taxonomic categories: metaphoric, iconic, and beat. Next, we calculated the agreement rate for each of the tasks to determine the degree of consistency between the proposed gestures for each task. Based on this data, we were able to define a gesture set that can be considered spontaneous and intuitive. Then, we evaluated each of the gestures based on its ease of execution and its goodness rate that was collected during the interviews.

Starting with the methodologies proposed by Kuehnel et. al. [5] that utilized a smartphone as the main gesture recognizing device, we wanted the spontaneity of the gestures to be the goal of our research, without the technological limitations. After defining this new spontaneous gesture set, we can focus on the technological aspect of recognizing the gestures. This opens new ideas for further exploration, such as using an EMG device to detect the muscle contractions when a user performs a finger snap, or using a smartwatch to detect a hand moving from left to right.

During the definition of the gesture-set, we noticed that some of the most intuitive gestures are very simple and that are used daily, such as moving the hand up or down. This opens up a new question of differentiating the intentional gestures from random hand movements. A possible solution could be adding an “initializing” gesture that triggers the detector, but this could lead to more complex and un-intuitive gestures, which is another question that requires further investigation.

# References

[1] Correia, Ana & Miranda, Leonardo & Hornung, Heiko. (2013). Gesture-Based Interaction in Domotic Environments: State of the Art and HCI Framework Inspired by the Diversity. 300-317. 10.1007/978-3-642-40480-1_19. [link]

[2] Vatavu, Radu-Daniel & Wobbrock, Jacob. (2015). Formalizing Agreement Analysis for Elicitation Studies. 1325-1334. 10.1145/2702123.2702223. [link]

[3] Wobbrock, Jacob & Morris, Meredith & Wilson, Andrew. (2009). User-defined gestures for surface computing. Proceedings of ACM CHI. 2009. 1083-1092. 10.1145/1518701.1518866. [link]

[4] Mcneill, David. (1994). Hand and Mind: What Gestures Reveal About Thought. Bibliovault OAI Repository, the University of Chicago Press. 27. 10.2307/1576015. [link]

[5] Kuehnel, Christine & Westerman, Tilo & Hemmer, Fabien & Kratz, Sven & Muller, Alexander & Moller Sebastian. (2011). I'm home: Defining and evaluating a gesture set for smart-home control [link]