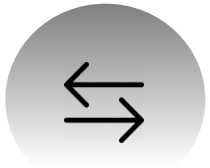
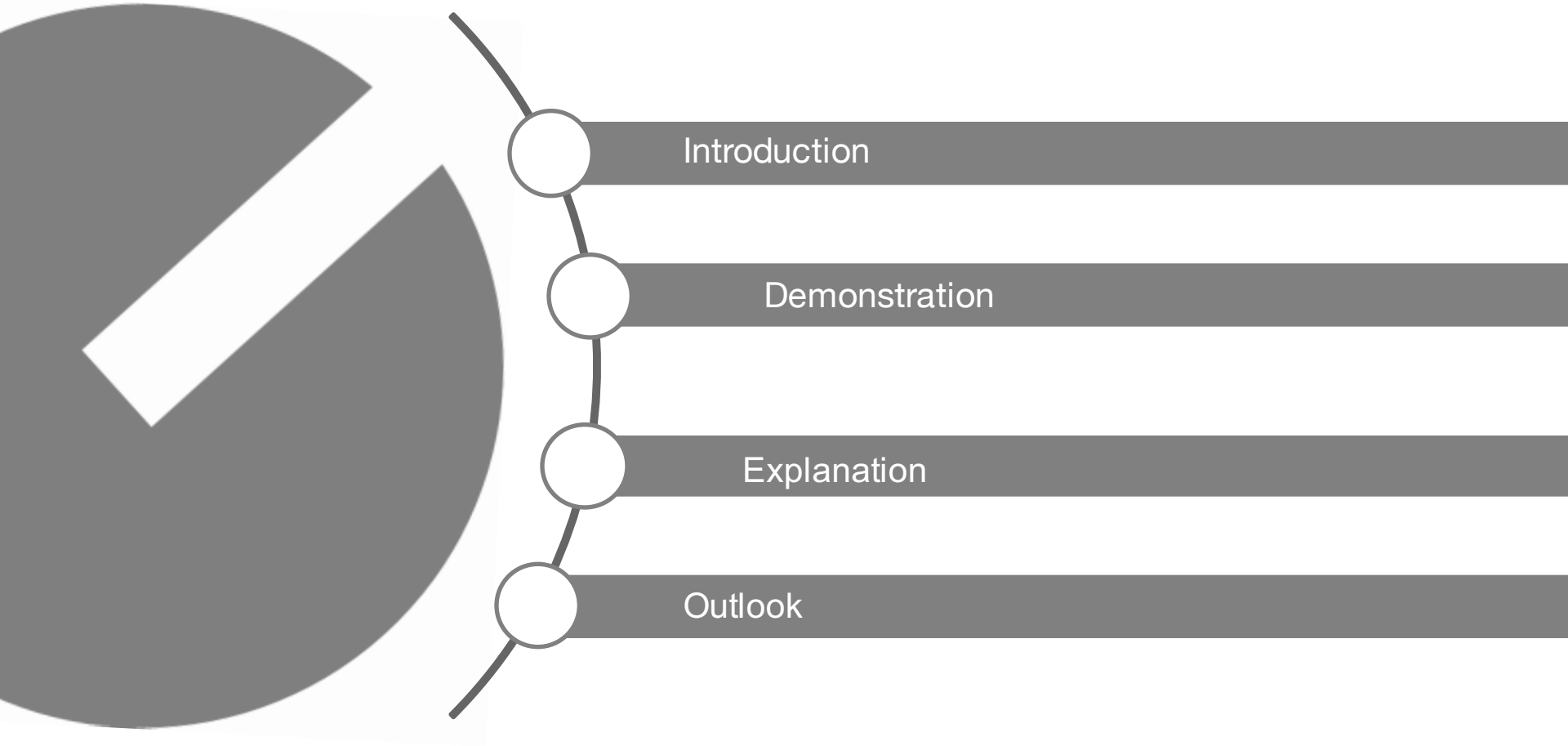




Beijing, 25th of May 2017

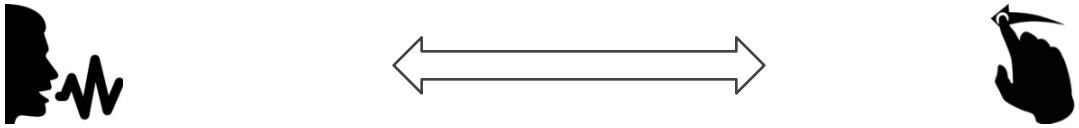
Gesture Tracking
meta.Me

Structure

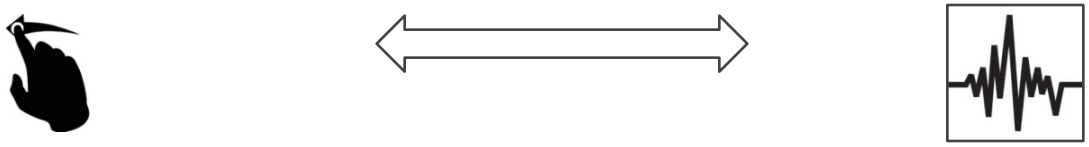


Recalling the Idea

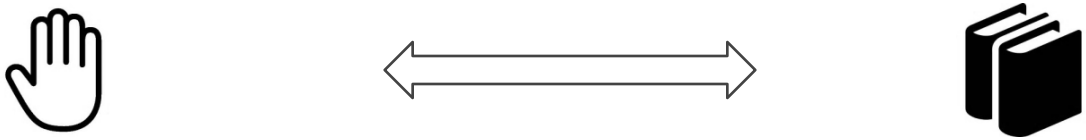
Bridging Voice Control and Gesture Sensing



Guiding Gestures in the air through Vibration Sensing

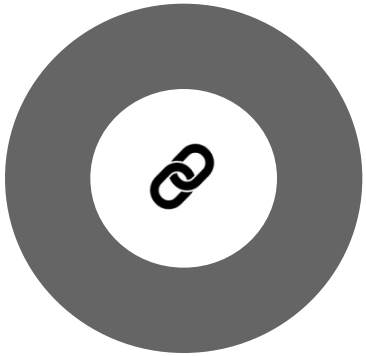


Combining practical and theoretical findings



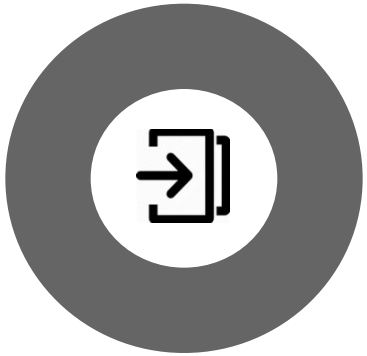
Recalling the Idea

A three Step Process:



Linking

Controlling



Leaving

Demonstration

meta Me

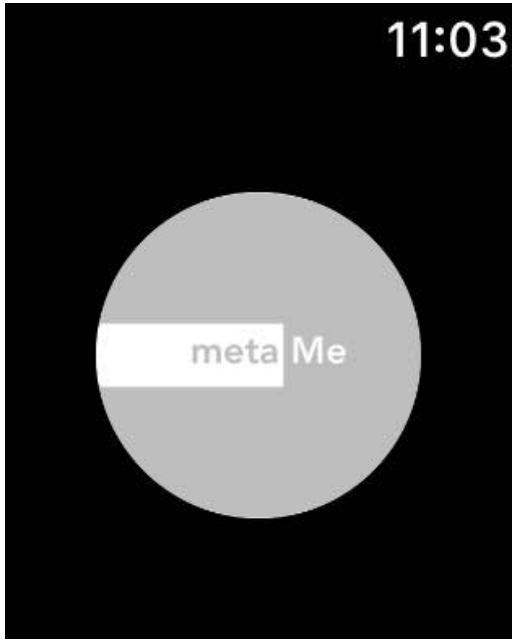
meta.Dial

Case 1
Music

Case 2
Light

Demonstration

Home Screen



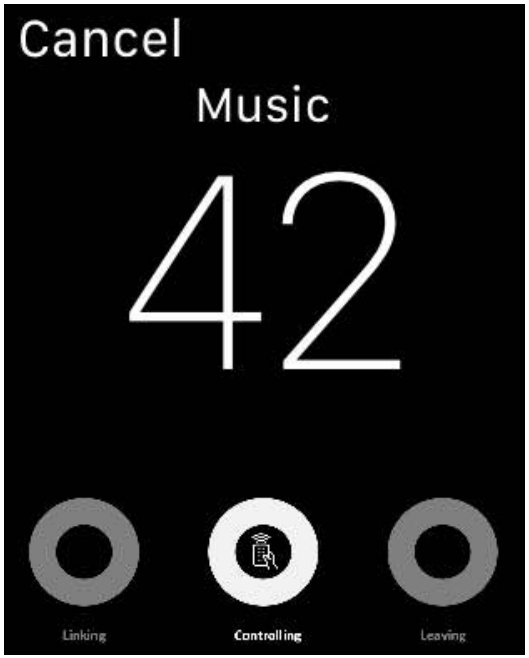
Demonstration

Voice Recognition

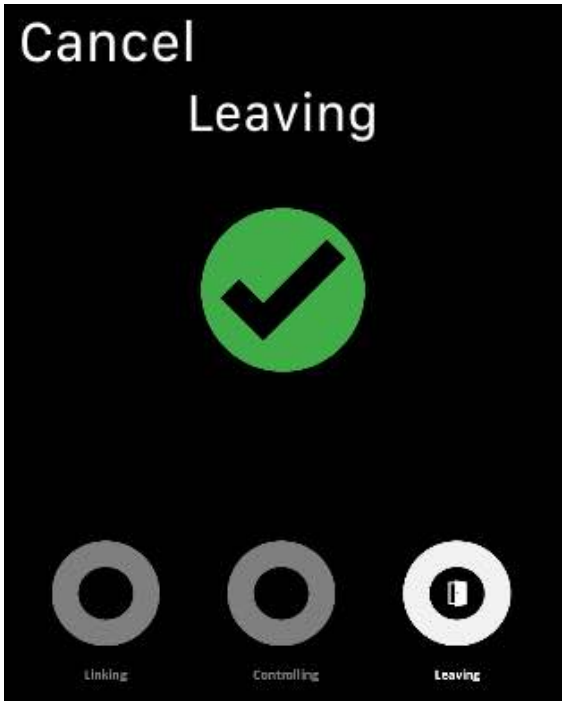
Linking Stage

Demonstration

Volume Adjustment

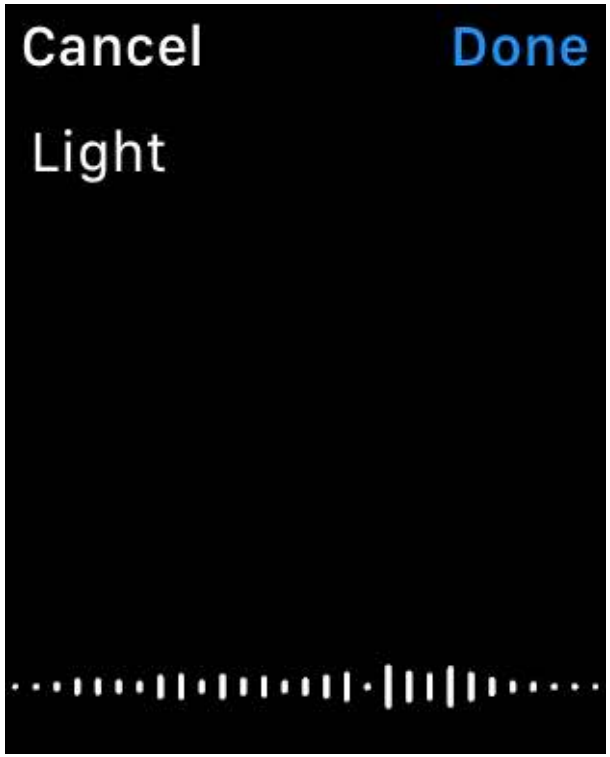


Leaving Stage



Demonstration

Voice Recognition

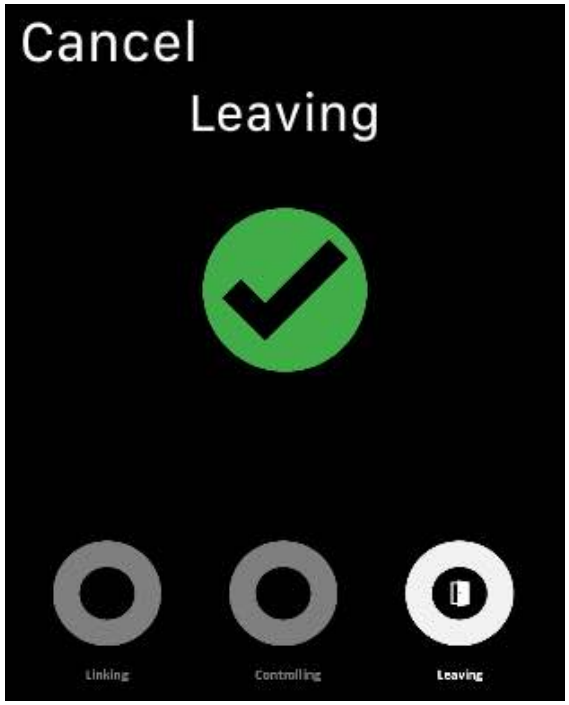


Linking Stage

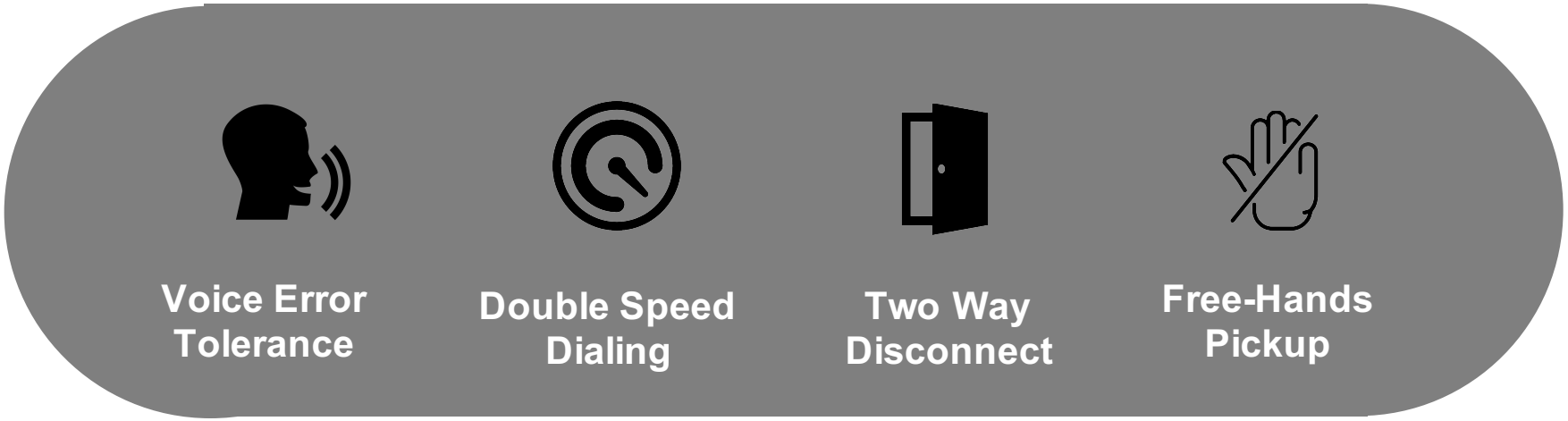
Demonstration

Light Adjustment

Leaving Stage



Add-Ons


Voice Error Tolerance
Double Speed Dialing
Two Way Disconnect
Free-Hands Pickup

Explanation

Issue:

Human Side

Solution:

Technology Side

Explanation

Problem:

Feedback

changing value

maximum border

Solution:

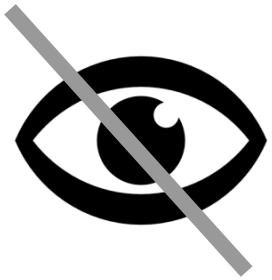
Vibrations

Guide Vibrations in the air

Explanation

Problem:

Feedback



… is value changing?

Solution:

Dial Vibration

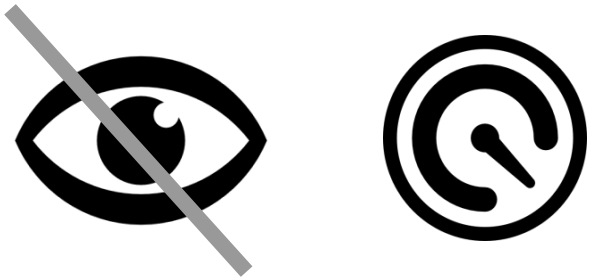
```
self.volume = (volume_start) + (self.roll - self.roll_start)

if abs(volume_vib - self.volume) > 5 {
    WKInterfaceDevice.current().play(.click)
    volume_vib = self.volume
}
```

Explanation

Problem:

Feedback



… is maximum reached?

Solution:

Border Vibration

```
if (self.volume) > 100 {
    self.volume = 100
    WKInterfaceDevice.current().play(.failure)
} else if (self.volume) < 0 {
    self.volume = 0
    WKInterfaceDevice.current().play(.failure)
}
```

Explanation

Problem:

Voice Input

Solution:

Multicase Voice Recognition

```
/*  if(self.data.lowercased() == "hi-fi"){
       self.volumeChange(value: Float(self.volume)/100)
  }*/

    if((self.data.lowercased().range(of: "hi")) != nil){
   self.volumeChange(value: Float(self.volume)/100)
    }

  if ((self.data.lowercased().range(of: "li")) != nil) {
       self.lightChange(value: Float(self.volume)/100)
  }
```

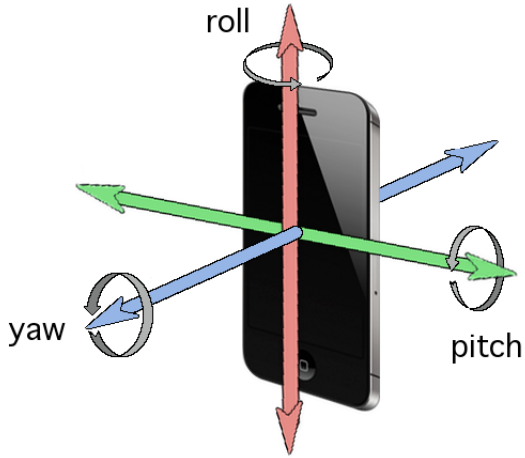
Explanation

Problem:

Gesture

Solution:

Roll-Yaw-Pitch



Explanation

Problem:

Gesture

Solution:

Accelerometer

```
func gesture () {
    var i = 0
    let volume_start = 50.0 //ATTENTION: needs to be read from hardware in final implementation
    var volume_vib = volume_start
    motionManager.accelerometerUpdateInterval = 0.1;
    motionManager.startAccelerometerUpdates(to: OperationQueue.current!) { (data, error) in
        if i == 0 {
            self.roll_start = atan2(-(data?.acceleration.y)!, (data?.acceleration.z)!) * 57.3
            if self.roll_start < 0 {
                self.roll_start = (180.0 - abs(self.roll_start)) * -1.0
            } else if self.roll_start >= 0 {
                self.roll_start = 180.0 - abs(self.roll_start)
            }
            WKInterfaceDevice.current().play(.start)
        }

        self.roll = atan2(-(data?.acceleration.y)!, (data?.acceleration.z)!) * 57.3
        if self.roll < 0 {
            self.roll = (180.0 - abs(self.roll)) * -1
        } else if self.roll >= 0 {
            self.roll = 180.0 - abs(self.roll)
        }
```

Explanation

Problem:

Volume

```
func volumeChange(value: Float) {
    let requestValues = ["command" : "volume", "level" : value] as [String : Any]
    let session = WCSession.default()

    session.sendMessage(requestValues as [String : AnyObject],
                        replyHandler: nil, errorHandler: { error in
                            print("error: \(error)")
    })
    // print(value)
}
```

Explanation

Problem:

Light

```
func startLight() {
    if WCSession.default().isReachable == true {

        let requestValues = ["command" : "startLight"]
        let session = WCSession.default()
        session.sendMessage(requestValues, replyHandler: nil, errorHandler: { error in
            print("error: \(error)")
        })
    }
```

Thank you for your attention.

Any Questions?

Table of Contents

Introduction and Mission Statement

meta.Me has set itself the goal of developing and implementing of a new dial metaphor for controlling devices in smart home. Throughout the new human computer interface, the input technologies of voice and gesture shall be closely intertwined. Therefore, we have come up with a low-level prototype emulating the basic functionality of the final product. Test users will be equipped with a smartwatch device and perform selected use cases. The individual use cases are all of differing difficulty spanning a wide field of application scenarios. The rationale behind the experiment is an conclusive understanding of user behaviour and preferences. Major use cases shall be performed and analyzed. This points out to the mission statement of the low-fi prototyping phase. Currently, the base frame of the meta.Dial is defined, however details are not yet specified. Ultimately aiming at an intuitive and user-friendly product, we hope to receive significant findings from the current development phase. Those include, how will feedback be transmitted, what kind of device should be worn, what kind of smart home devices can be conveniently controlled, to name just a few. We particularly focus on the intricacies of user perception. The purpose of the project is to provide users with a convenient way of adjusting and manipulating smart home devices. Furthermore, the meta.Me teams targets to collect pioneering experience in combining the benefits of voice and gesture control. Another leading aspect is research in natural vibration feedback for guiding gestures in the air. The gain from the project will not only be a final end product, but also theoretic findings and guidelines.

Prototype

A straightforward and effective tool, is mostly characterized by simplicity. Consequently, our low-fi experiment setup also exhibits this specification. When using the final product, the applicant will solely be equipped with a smartwatch controlling different smart home devices. The smartwatch however, has to have a few technological features. Those features shall be assumed in the prototype imitating the final experience.

The main pieces of functionality are voice recognition, vibration and gesture sensing. The entire use case can be well structured into three parts. The “Linking”, the “Controlling” and the “Leaving”. Irrespective of the use case, the applicant can always refer to this structure.
As for the linking, the different smart home devices will be addressed by voice. In particular, the respective device will be called, putting the smartwatch closer to one’s mouth. The linking phase terminates when the gadget is addressed, haptically and visually underlined by the watch. The watch will give the first of the three conclusion vibrations. Also, the linked gadget will be shown on the display.

The next step is the controlling phase. For adjusting the smart home devices, a conclusive interplay of haptic sensoring will be generated. Hence, the watch will give a second vibrating signal when the user can start to make its movement. The hand movement in the air emulates the turning of a real existing dial knob. Thanks to this movement, the project is called meta.Dial. Since hand gestures can be hard to carry out without having feedback, our team has worked hard on finding a natural and rational solution. The gestures will be supported and guided by varying vibrations of the watch outlining the range of input. For instance, different strengths will denote continuous ranges. Whereas for discrete ranges, the watch will generate “popping” or “flipping” vibration effects. This second phase can be ended through stopping the movement and holding still for a short pre-defined time. The most convenient duration of an optimal time lag between every phase shall be found through experimenting and deeper research.

Indicated by the third phase conclusion vibration, the controlling phase will make room for the leaving phase. During this phase, only visual feedback is provided. Now, the display of the watch will show the adjusted state of the smart home device and will be giving a confirmation of the executed action. In short, these are the main interaction concepts.

The central part of our prototype is a *Huawei Watch*, which shows the camera image captured by the connected Android phone using the app *PixtoCam*. The watch UI is displayed on a laptop, captured by the phone and streamed in real time to the watch, so our team can control the UI from the laptop according to the actions of the user. Additionally, we are able to generate different vibration patterns on the watch remotely from the phone using the app *Feel the Wear* to give the user an approximation of the vibration feedback we conceived for the product.

Methodology

As we described in our contextual inquiry, we want to put a focus on students as our target user group. Therefore, we looked for participants in Coffee Paradiso in the first floor of Zijing Building 21, Tsinghua University. We asked three persons to participate in our user study. Our test group included two females (27, China, and 18, France) and one male (40, Canada). Our participants had not been exposed to our project beforehand in any way.

We used a laptop to display the watch UI screens and installed a phone on a tripod to record the laptop screen, the image captured by the phone is streamed to the smartwatch worn by the user, as described before. Additionally, we used a second laptop to simulate the device that is controlled in each task. This laptop is remotely controlled by a second Android phone according to the user actions using the app *Unified Remote*. The user-facing parts of this setup are only the watch and the second laptop, the other parts (laptop showing UI, phone recording and streaming UI and generating vibrations, phone used to control second laptop) are “behind the scenes”.

Our three tasks cover a wide range of different devices in the smart home and show the versatility of meta.Dial. The first task is to simply switch on the light. This is a task with only two states, light on and off. The result is instantly visible to the user. For this task, our second laptop shows the image of a darkened living room. As soon as the user performs the dial gesture, this image will be remotely switched to the image of the fully lit living room. The vibration feedback given during the dialing consists of only one short vibration when the user dials pass the border between “on” and “off”. Note that the other vibration feedbacks indicating successful linking and beginning and end of dialing stay the same throughout all tasks.

The second task is to adjust the volume of the HiFi system. While the result of the adjustment is still instantly apparent to the user through the change of volume, this task involves a wide range of states instead of only on and off. During this task, the second laptop shows the image of a hifi system and plays a song in the background. The volume of the laptop is remotely controlled by the phone according to the user action. The vibration feedback during dialing involves a series of very short vibrations, one vibration for every few degrees the user rotates his/her hand.

The last task is to control the room temperature by setting the air conditioner. As in the second task, this involves a wide range of possible states, but the result is not instantly perceivable, since the room temperature will not change instantly. For this task, the second laptop can not simulate the result of the user action and is therefore omitted. The vibration feedback is the same as in the previous task (this is only due to limitations of our prototype setup, our team has different feedbacks for these two tasks in mind, though the details are not worked out yet). Both haptic and visual feedback by the watch are given in the previous tasks, too, but they do not play a key role in those tasks, as opposed to this last task. Since there is no direct feedback from the controlled device in this task, the vibration feedback

becomes much more important than in the first two tasks. The visual feedback given by the watch UI at the end of the dialing, which displays the state just set by the user, in this case the new temperature, is also vital in this task to give the user a sense of assurance about his/her action.

We went through this procedure by assigning clear roles for our team members. Tony was responsible for operating the smartwatch UI, i.e. clicking through the slides for the UI on the laptop at the right time and generating the appropriately timed vibration feedbacks through the app on the recording phone. Yura controlled the second laptop, which simulates the controlled devices, via the *Unified Remote* app on a second phone. This means for the first task, the image of the dark room had to be switched to the bright room at the right time. In the second task, the volume of the second laptop had to be adjusted appropriately. Markus and Nik observed the participants carefully and logged the critical incidents. The participants were recorded by video camera as backup for later review.

Our team members have a clear idea of how the system is supposed to work, therefore we may not be able to unreservedly assess the intuitiveness of the whole procedure, especially what errors users might make. Therefore, a big focus during the user study is to observe the errors that happen, so that we can later adjust our design accordingly. It is also of interest what kind of voice commands the participants use to trigger the desired device, since one of our goals with meta.Dial is to overcome the need to remember complex voice commands and syntax traditionally present in voice controlled smart home systems. Another important aspect is to observe which directions (clockwise or counterclockwise) the users are dialing in the different tasks. Last but not least, we hope to be able to draw conclusions about the overall intuitiveness and usefulness of meta.Dial.

Results

The common result from three participants is they were satisfied with the experiment and seemed very excited about how they control the smart home. After we give instruction about how to do it, understand quickly and tend to do pretty accurate as we imagine. Users all admit that they like to use voice and gesture input and also said it was easy to learn. Moreover, give us comment that vibration was helpful for controlling. Our participants were all agree that if the system is implemented, they would like to use in the smart home. But, we also observed that there were some of the problems that we need to work on.
The first common problem was they were confused to control the device at the first time so that we need to do the experiment again. After speech input, they were tend to hesitate the moment to do the gesture motion.Second common problem was when participants start to do the dial gesture, they were repeat the gesture or just do the one side dial gesture. Third, some participant use different terms for the speech input.

Our first participant was 27 female language student. She is Chinese so there was little bit of conversation issue. When she learn our prototype she was bit confused about input so directly use natural speech input like “Turn on the light”. And she was also confused how to do the dial gesture so repeat the gesture few times. But after the first task, she was fine to do with other two tasks. For the experiment survey she gave us positive points for the questions and also gave us comment that it was easy to operate with few words and simple gesture. In addition, for the negative comment she said that the dial ranges were not enough for the light and she thinks that more ranges are better for the user experience.

Second participant was female language student whose age is 18. She is from France and she was catch pretty fast about the instruction compare to the first participant. Basically, She was very good with all 3 tasks but also has problems with gesture that repeat two or three times. For the experiment survey she gave all 100% positive point. Her comment was useful and also gave us comment that speech recognition won’t be work all the time due to the situation that people can’t talk.

Last participant was computer engineering male. His age is 40 and he is from Canada. He followed very well with our instruction and did the experiment very smoothly. He completed the 3 tasks quickly, but for the second task which is Hi-Fi system, he also control with nature language like “Raise the volume”. He gave us most critical comment like vibration wasn’t strong enough and need more vibration level. But overall he was satisfied with the system and agree that remote control is good.

Discussion

First, we need to figure out the direction of the dial gesture that which direction is comfortable and user friendly. From our prototype result we learned that users were generally tend to give only one direction for the gesture. Therefore, we need to find out which direction is suitable with the ranges. Furthermore, in the experiment we didn't reveal yet, but for the implement, our team need to work on the problem of wrong direction feedback. When users do the dial gesture, if they operate wrong direction, they might only can figure out this after they check the smartwatch screen. So, better have feedback for the wrong direction gesture.

Second is late feedback problem of Task 3 which is temperature. There were issues with gesture that participants repeat it, but it's because for the prototype that they didn't get direct feedback, so we expect this problem will be generally solved after implementation. However for the temperature, there is a still problem of delay response like when the users adjust the temperature, they might tend to repeat the gesture. So we need to have a solution of this.

Third, define the terms of speech input. One of our participant was using temperature term as "air conditioner" and other used "air condition". In addition, for Hi-Fi system participants said "stereo" or "volume". Therefore, we need to find natural terms for users to fit the speech input. One more issue about speech input is that during the experiment, participants made mistake like "Turn on the light". Users sometimes say natural language as speech phrase. So, users need to learn properly for the speech terms.

Last, better vibration design is needed. Our team got a feedback about vibration from one of our participant. He gave us comment that vibration is not strong enough and also more levels for vibration is necessary. Thus, the intense of vibration and level of it should be figure out.

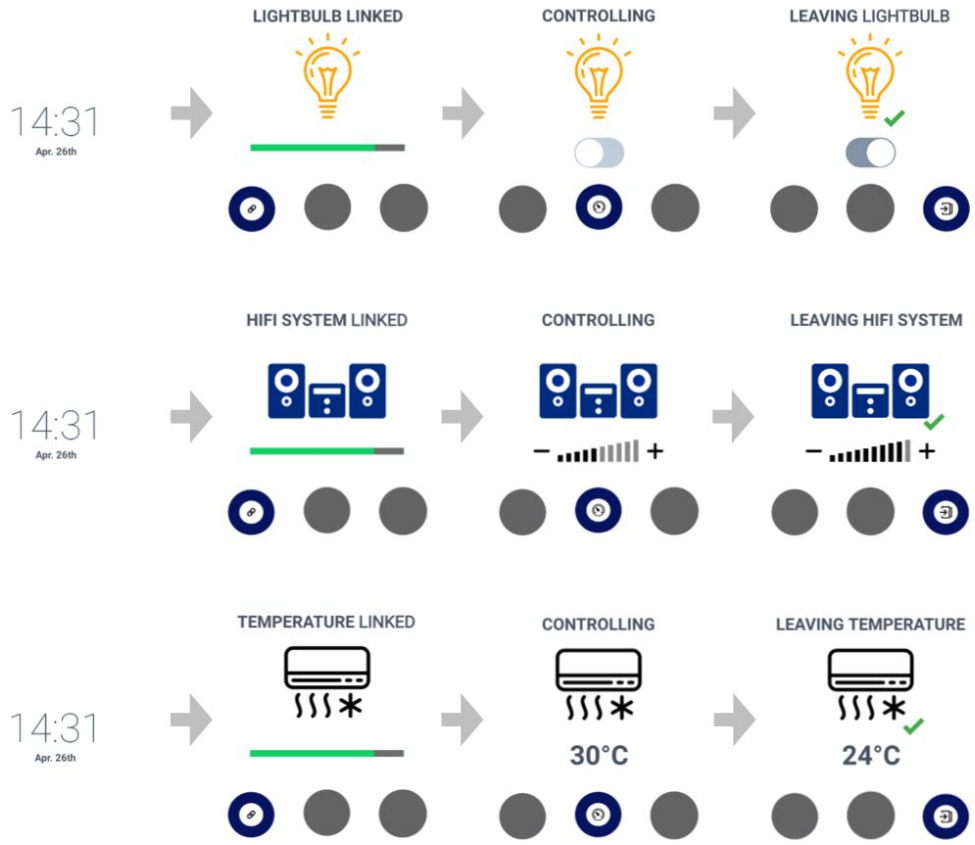
In order to evaluate the severity ranking, we used the idea of diving the severity in three factors and each factor is rated on a scale from 1 to 10:The frequency, that is to say, how often an event happens to the user (10 means very frequent), the impact of the problem, that is to say, how difficult is it to overcome (10 means difficult), and the persistence meaning if a problem once it is known is easy to overcome or if overcoming is difficult every time (10 means overcoming is very difficult every time).

	(F)requency	(I)mpact	(P)ersistence	Severity (F x I x P)
Wrong direction	3	5	6	90
Gesture repetition for temperature	2	3	5	30
Speech input terms	5	7	2	70
Noticeability of vibration feedback	9	6	3	162

The results from the severity ranking should shortly be commented:
Generally, the problems evaluated in this comparison have a low severity, given the maximum value achievable in this comparison would be 1000. From the severity rating, one can see that the noticeability of vibration feedback has highest relative importance; this seems intuitive given the state of the system is not present to the user without looking on the display when the vibrational feedback is unnoticed. For the speech input term issue, giving the learning behaviour of the user, the persistence reduced the influence of the impact and frequency, yet, in order not to frustrate new customers, this topic needs to be faced.

Appendix

Appendix A: GUI Screenshots



Linking **Controlling** **Leaving**

Appendix B: User study script

Part 1: Introduction
We are doing a project for our Human Computer Interaction Technology Class. The general topic is smart home. Have you ever heard of smart home before? There are many smart home devices which can be remotely controlled, companies right now tend to realise it via a smartphone app or voice control. We want to replace it with a combination of voice and gesture control.

Part 2: Demo
So the user wears this smartwatch to control the different smart home devices. The whole interaction consists of three steps. The first step is to select and trigger the desired device through a simple voice command. Imagine you are in the kitchen, your hands are dirty, and you want to turn off the stove without touching anything. So the first step in this case would be a command like “Stove!” (*Raising the watch closer to the mouth, saying “Stove!”*). A vibration as well as the screen on the watch will tell you that the stove was successfully selected. The second step is to do the gesture. We rely on a simple dial gesture (*demonstrate gesture*) for all our control tasks. The idea is that your turning an imaginary dial/knob to control the device. So like you would turn the dial on a traditional stove, you do here, too, only in the air and remotely. So after a short amount of time, a second vibration and the watch screen indicate that you can start the gesture to turn off the stove (*demonstrate the gesture*). The last step is a short confirmation of your action, again through vibration and on the display. So the in actual speed, it would be like this: (*Raising the watch closer to the mouth, saying “Stove!”. Demonstrate the gesture. Glance at the watch as if checking the final result.*)

Part 3: Task instructions
So we have three different tasks for you that you are supposed to solve with this system. The first task is, you are in this living room (*point to the laptop with the image of the dark living room*) and you want to turn on the light. (*Let the user do the task*). The second task is, you are listening to music on your hifi system (*point to the laptop with the image of the hifi system, let music play*) and you want to raise the volume. (*Let the user do the task*). Now the last task. You are in the room, and it is much too warm inside, so you want to turn down the room temperature. (*Let the user do the task*).

Appendix C: Empty usability questionnaire

Part 1: Agree/Disagree questions

I agree on the following statements ...	100%				0 %
I am able to efficiently complete my work using this system					
I feel comfortable using this system					
It was easy to learn to use this system					
The interaction with my smart home devices is pleasant					
I like using gesture to control the system					
I like using voice to control the system					
The vibrations helped me in the interaction					
Overall, I am satisfied with this system					

Part 2: Open questions
What was positive? What was negative?
The questionaire partially used questions from http://garyperlman.com/quest/quest.cgi, checked 26.04.2017

Appendix D: Log of critical incidents

Participant 1 – Task 1:

- Did not get the concept for the voice command correctly in the beginning, said “turn the light on” instead of only “light”
- Did not raise the arm for the voice command, probably due to our comment to start in a natural position, meaning starting with arm down, probably misunderstood this
+ Seemed to be pleasantly surprised by the result

Participant 1 – Task 2:

- Did not raise the arm for the voice command, probably due to our comment to start in a natural position, meaning starting with arm down, probably misunderstood this
- Instead of one smooth gesture in one direction, repeated the gesture multiple times

Participant 1 – Task 3:

- Did not raise the arm for the voice command, probably due to our comment to start in a natural position, meaning starting with arm down, probably misunderstood this
- Instead of one smooth gesture in one direction, repeated the gesture multiple times
+ Seemed to be pleasantly surprised by the result
+ Very excited by the whole experience (“So cool”, “best thing”)

Participant 2 – Task 1:

+ Seemed to be pleasantly surprised by the watch “picking up” the voice command
- Instead of one smooth gesture in one direction, repeated the gesture multiple times
+ “Oh, cool” when the light “actually” turned on

Participant 2 – Task 2:

+ Seemed to be delighted about the volume
- Made the dial gesture in the wrong direction
+ “That’s good!”

Participant 2 – Task 3:

- Wasn’t sure about the right voice command
- Instead of one smooth gesture in one direction, repeated the gesture multiple times

Participant 3 – Task 1:

- Did a raise hand gesture instead of the dial gesture in the first try

Participant 3 – Task 2:

No critical incident

Participant 3 – Task 3:

- Did not check the watch in the end, therefore not sure whether the action was successfully completed

Appendix E: Raw data usability questionnaire

Task was just turn on and turn off, so have more ranges will be give better experience.

Participant 1:

Part 1: Agree/Disagree questions

I agree on the following statements ...	100%				0 %
I am able to efficiently complete my work using this system		X			
I feel comfortable using this system	X				
It was easy to learn to use this system	X				
The interaction with my smart home devices is pleasant	X				
I like using gesture to control the system		X			
I like using voice to control the system	X				
The vibrations helped me in the interaction	X				
Overall, I am satisfied with this system	X				

Part 2: Open questions

What was positive?

- Easy to operate with few words and simple gesture

What was negative?

- Task was just turn on and turn off, so have more range will be give better experience

Participant 2

Part 1: Agree/Disagree questions

I agree on the following statements ...	100%				0 %
I am able to efficiently complete my work using this system	X				
I feel comfortable using this system	X				
It was easy to learn to use this system	X				
The interaction with my smart home devices is pleasant	X				
I like using gesture to control the system	X				
I like using voice to control the system	X				
The vibrations helped me in the interaction	X				
Overall, I am satisfied with this system	X				

Part 2: Open questions

What was positive?

- Useful

What was negative?

- If you have a situation where you can’t speak, speech recognition wont work

Participant 3:

Part 1: Agree/Disagree questions

I agree on the following statements ...	100%				0 %
I am able to efficiently complete my work using this system		X			
I feel comfortable using this system		X			
It was easy to learn to use this system		X			
The interaction with my smart home devices is pleasant		X			
I like using gesture to control the system	X				
I like using voice to control the system	X				
The vibrations helped me in the interaction			X		
Overall, I am satisfied with this system		X			

Part 2: Open questions

What was positive?

- Remote interaction

What was negative?

- Vibration level wasn't strong enough.
- Vibration level could be stronger