


Explore Gesture interaction Support Explaining ideas

The word "Gesture interaction" is highlighted in a light blue box. Two hands are shown: one on the left pointing towards the box, and one on the right pointing away from it.

Explore how hand moving gesture interaction can support user explaining ideas among a group of audiences during the remote meeting or physical presentation.

Ziyue Wang

Supervisors: Mosse Sjaastad, Lars Marcus Vedeler

Prototype

<https://gesturemeeting.netlify.app>

Abstract

This is an exploratory diploma project, it explored the different possibilities of using gesture recognition technology (to be more specific, recognize people's hands through camera by machine learning) to support people explaining ideas in remote meeting.

By combing sketching, programming and testing, different ideas were generated and those ideas are relevant to different aspects of meeting. Besides those ideas, I also explored some basic problems of how people can interact with machine/computer through hands.

In the final refined sketch, people can use gestures to create shapes, make lists, draw lines or move those elements in a transparent "blackboard." By those elements, they can explain their ideas better. If the user is sharing slides, both the presenter and audiences can also move their videos through gesture. However, the refined sketch is not the only solution, each function in the refined sketch can lead to more possibilities.

As an exploratory diploma, the goal of this project is showing different possibilities, along with the potential solutions and insights for interacting with machine through hands, this project builds a foundation and people can go further from here.

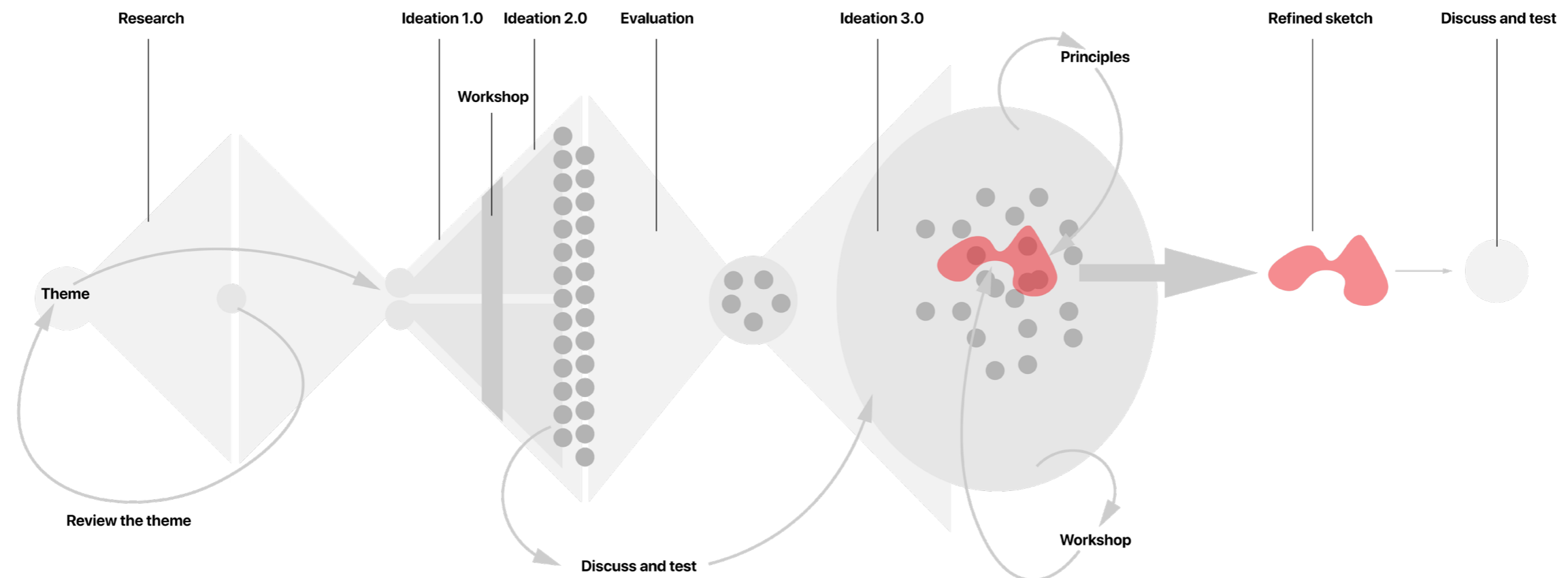
Why

Gestures during the communication are ubiquitous across culture, it can illustrate pictures and thoughts which are hard to express by language alone.

With the development of gesture recognition, there are more and more exploration of how to apply this technology in our daily life, but there is barely discussion of how to combine gesture recognition with our natural gestures to support us explaining ideas and thoughts, especially under the situation in which remote meeting and presenting ideas become more and more popular. Therefore, I want to explore how gesture interaction can help user to explain his ideas and thoughts among a group of audiences during the remote meeting or physical presentation.

My process

This project started with a broad theme, I first did desktop research of human's natural gestures, research and trying different gesture recognition technology, desktop research of meetings. After research, I reviewed my theme, narrow it down and came up with 2 paths to explore the project. Along with those 2 paths, I did first round of ideation "Ideation 1.0", to generate more ideas, I held a workshop after "Ideation 1.0", to get more inspiration, then I did second round of ideation "Ideation 2.0", After that, I discussed those ideas and tested some prototypes with the experts, regular people and my supervisors. As I can't explore all aspects at same time, I evaluated those ideas and choose a group of ideas to go further. Then I did the third round of ideation "Ideation 3.0". In "Ideation 3.0" I found another research was needed, so I held another workshop. Besides ideas, in "Ideation 3.0" I also came up with some principles based on my insights, with those principles and the results from workshop, I integrate part of ideas in "Ideation 3.0" and came up with the refined sketch. Then I discussed the refined sketch and tested the prototype with experts in the remote meeting field as well as regular people.



Catalogue

Research

- Research about natural gestures
- Research about gesture recognition technology
- Research about meeting
- Gestures in remote meeting & physical presentation

Review original theme

- Unclear part
- 2 Paths to explore

Ideation 1.0

Ideation 2.0

- Workshop
- More ideas

Testing and discussion 1.0

Evaluation

Ideation 3.0

Refined sketch

- Refined sketch
- Overview
- Setting gesture
- Modeless mode
- Function with mode

Principles

Testing and discussion 2.0

Reflections

Research

The project is about exploring how can we use gesture recognition this technology to recognize our nature gestures during the communication, obviously the research about people's natural gestures, gesture recognition technology and meeting should be conducted.

Research about natural gestures

- Desktop research
- Different type of gestures
- Function of different gestures
- Why gestures are important

Research about gesture recognition technology

- Desktop research and trying

Research about meeting

- Desktop research
- Different types of meeting
- Important aspects about meeting

Gestures in remote meeting & physical presentation

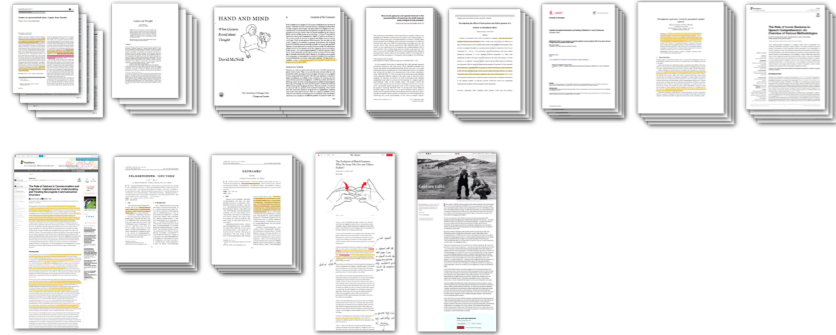
- Workshop
- Comparison and analysis
- Some interesting findings

If there is an QR code beside the image, please scan it or click the hyperlink under the image to watch the video, because gestures are dynamic, it's hard to explain with just words and static image.

Research about nature gestures

Desktop research

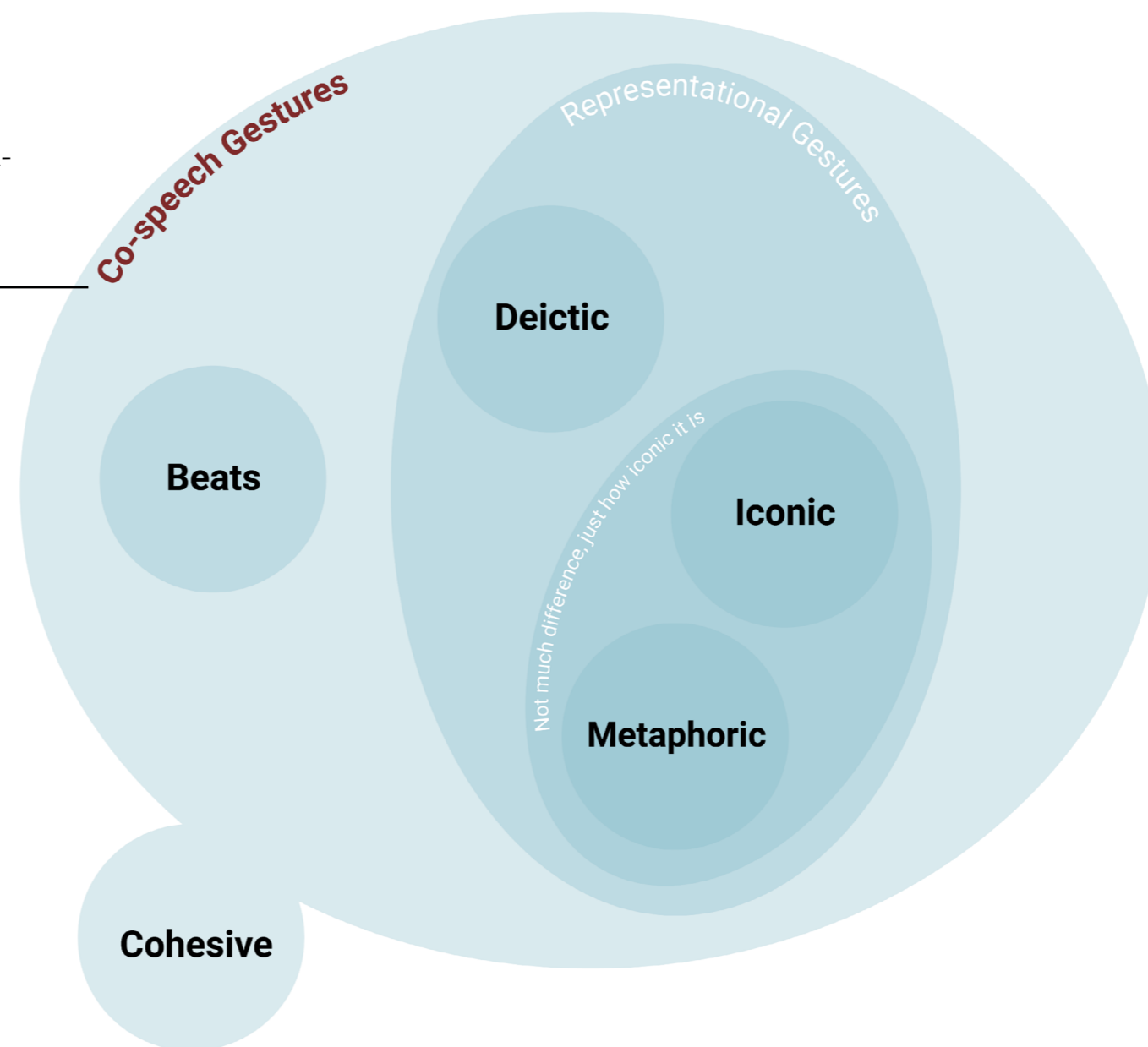
There are already many professional academic research about human's natural gestures, so I started the desktop research about gestures by reading papers and articles around this field.



Different type gestures

After the desktop research, I mapped out the different types of gestures. According to the desktop research, there are two categories of gestures, co-speech gestures and emblem gesture, within co-speech gesture category, there are 5 types of gesture: Beats, deictic, iconic, metaphoric and cohesive gestures. Co-speech gestures can only make sense when accompanied by speech while emblem gestures can represent the meaning without talking, such as thumb up. Interestingly, co-speech gestures have similarities across cultures but emblem gestures are different according to culture.

- Only make sense when accompanied by speech
- Similarities across cultures



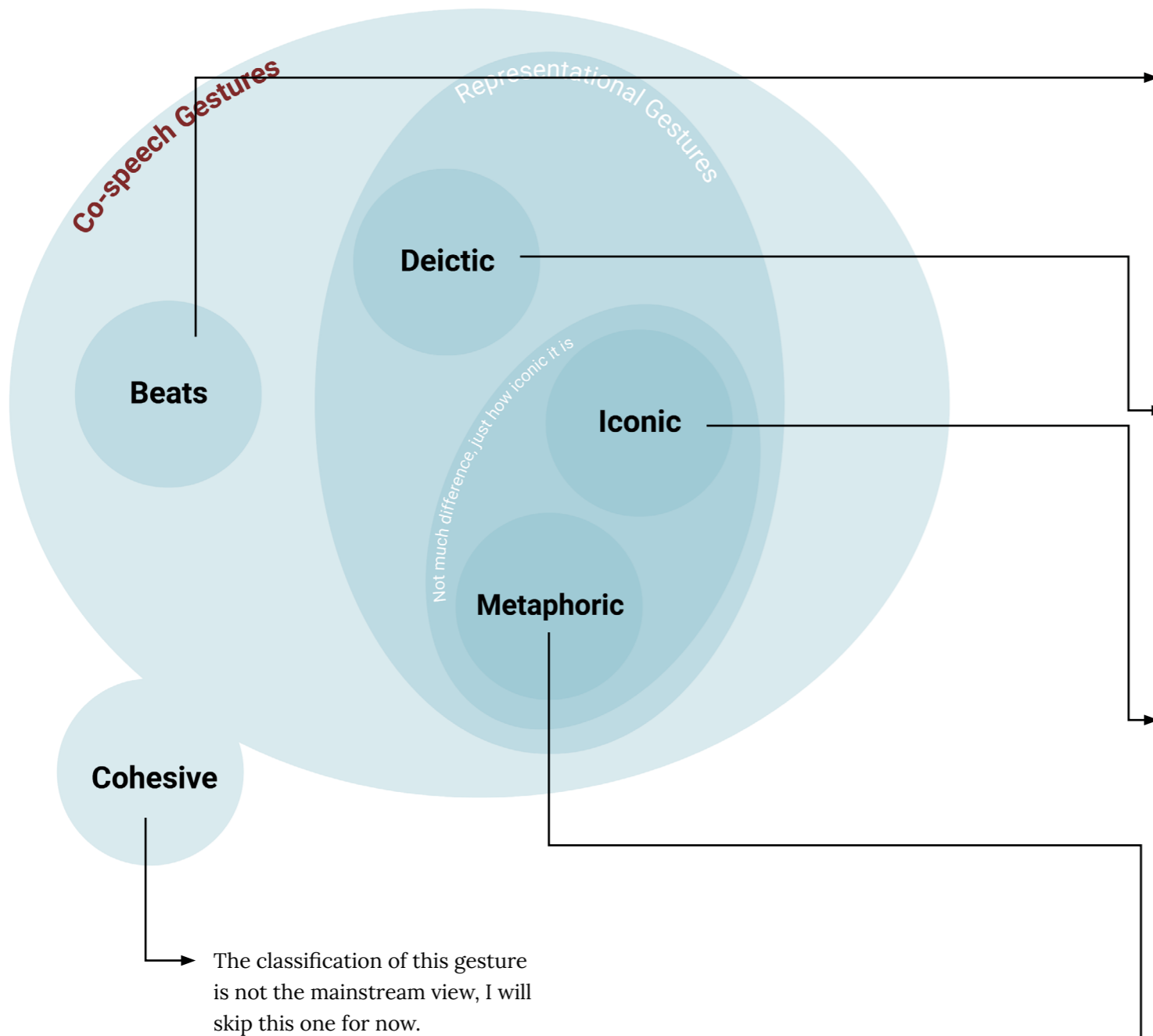
E.g. Thumb up

- Meaningful without talking
- Tend to be culture specific

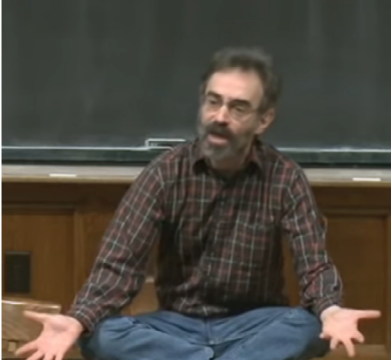

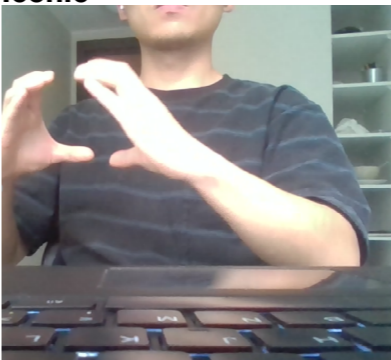
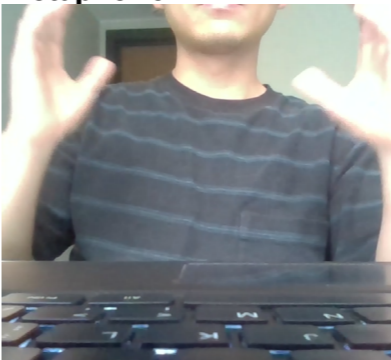
Research about nature gestures

Function of different gestures

The function of emblem gesture is obvious—represent the meaning behind the gesture. From the research, I also summarized the functions of different co-speech gestures.



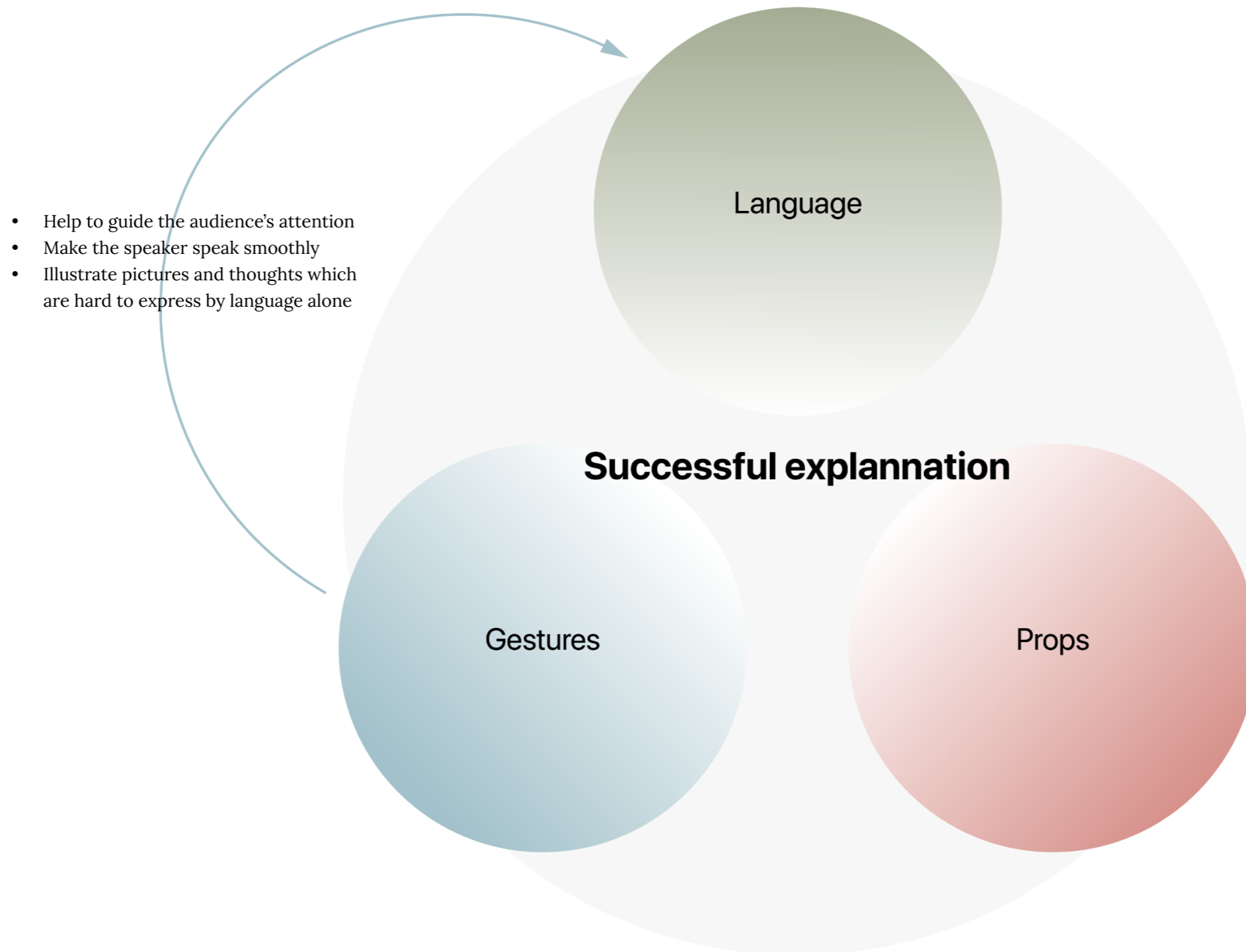
The classification of this gesture is not the mainstream view, I will skip this one for now.

	What	Functions
Beats  Click to watch the video	<p>Hand movements that occur in the rhythm of the language</p> <ul style="list-style-type: none"> Gestures accompanying phrases like “first, second”. Up-and-down or back-and-forth hand movements. <p>Example Speak about the important words of the sentence, there will be hands up and down movements.</p>	<ul style="list-style-type: none"> Uncover the structure of speech or sentence Draw attention Show rhythm
Deictic  Click to watch the video	<p>Concrete or abstract pointing movements.</p> <p>Example “back in my hometown” point backwards, or point at the slides.</p>	<p>Direct the attention directly to visual information</p>
Iconic  Click to watch the video	<p>Signals that convey meaning in both physical form and manner that co-occur with verbal meaning. More about concrete idea.</p> <p>Example Using hands to represent the shape and size when describing the lens of the camera, using hands to show the action of throwing a ball when talking about it.</p>	<ul style="list-style-type: none"> Information on actions (kinet-graph) Shape/size of an object (picto-graph) Spatial relationship between two objects (spatial movement)
Metaphoric  Click to watch the video	<p>Illustrated content does not refer to something physically present, but to a more abstract idea.</p> <p>Example During saying “I have a big idea”, people’s hands open up to transfer the abstract concept “big idea” to the physical size of big.</p>	<ul style="list-style-type: none"> Turns abstract ideas & structures into the visual and the embodied

Research about nature gestures

Why gestures are important

Why gestures are important to our communication? Briefly speaking, a successful explanation is a symphony of language, props and gestures. Gestures help to guide the audience's attention, make the speaker speak smoothly, illustrate pictures and thoughts which are hard to express by language alone.



Research about gesture recognition

Desktop research and trying

Not just doing the desktop research of different gesture recognition technology, I also tried my best to experience some of them: Google Soli, MediaPipe, Myo Armband. Some technology is not valid anymore (project Prague), some of them is already out of date (Myo armband).

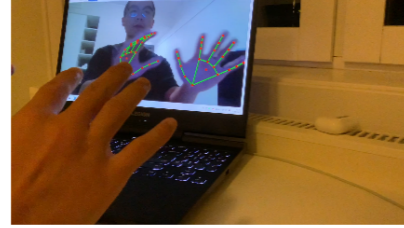


Google Soli

Device: Radar

Reflections:

- Not open to 3rd party
- Can detect detail finger movements
- Limited gestures
- Don't require camera

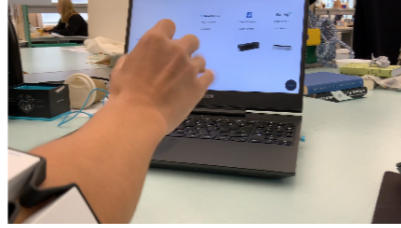


MediaPipe

Device: Camera

Reflections:

- Only need normal camera
- Easy to access
- Get coordinates of hand joints

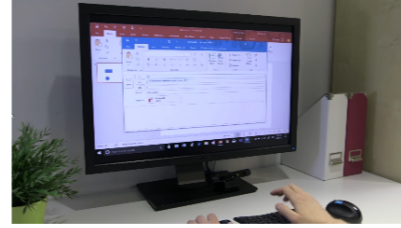


Myo Armband

Device: Armband

Reflections:

- Need additional device
- Not that accurate



Project Prague

Device: Depth camera

Reflections:

- Not valid now

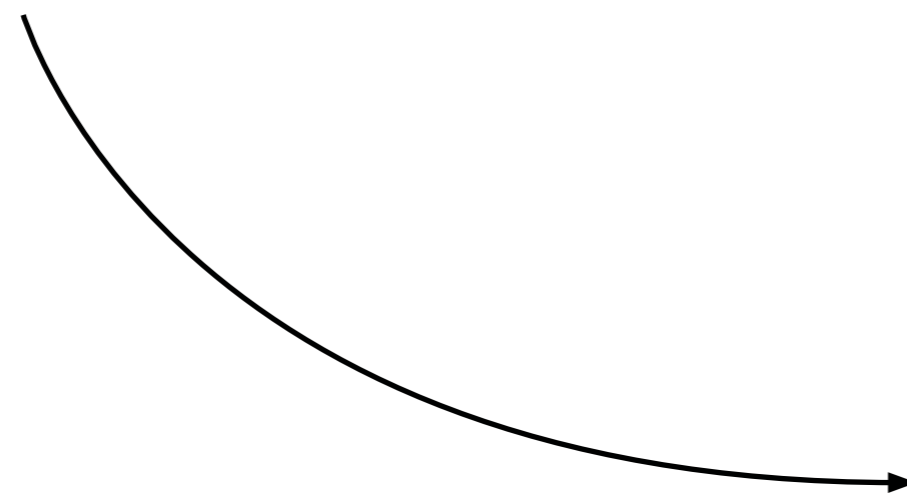


Leap Motion

Device: Leap Motion

Reflections:

- Need additional device
- Need to perform on top of the device



MediaPipe

After the research of those different gesture recognition technology, I decided to choose MediaPipe as technical basis of my project, because it does not require additional device, easy to access (has a JavaScript library), furthermore, it can get the coordinates of the hand joints rather than provide some certain gesture, that leaves more imagination.

Research about meeting

Desktop research

As an exploratory project under the context of meeting, I conducted a general research of meeting to build general understanding of meeting.



Part of the articles I read

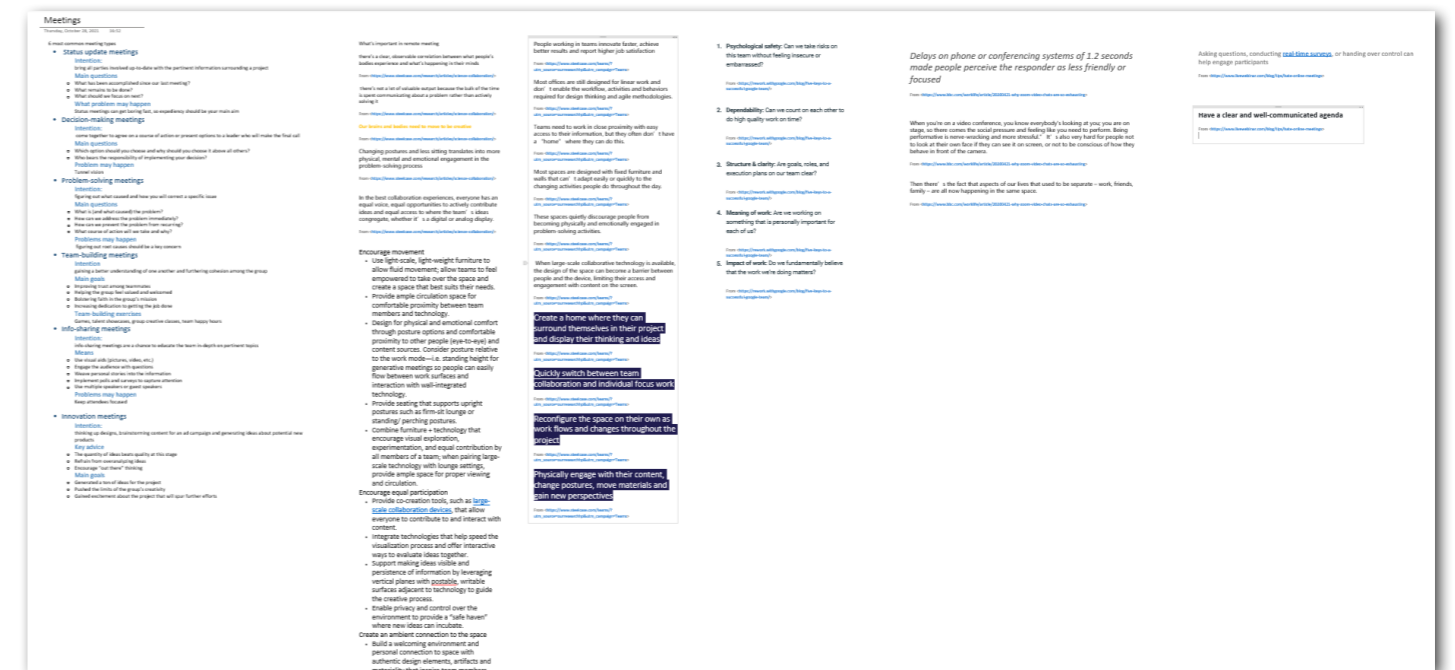
Different types of meeting

After the research I found 6 different types of meeting, summarized the intention, main goals, problems people may meet in those meetings.

<p>Status update meetings</p> <p>Intention: bring all parties involved up-to-date with the pertinent information surrounding a project</p> <p>Main questions:</p> <ul style="list-style-type: none"> • What has been accomplished since our last meeting? • What remains to be done? • What should we focus on next? <p>problems: Status meetings can get boring fast, so expediency should be your main aim</p>	<p>Decision-making meetings</p> <p>Intention: come together to agree on a course of action or present options to a leader who will make the final call</p> <p>Main questions:</p> <ul style="list-style-type: none"> • Which option should you choose and why should you choose it above all others? • Who bears the responsibility of implementing your decision? <p>problems: Tunnel vision</p>	<p>Problem-solving meetings</p> <p>Intention: figuring out what caused and how you will correct a specific issue</p> <p>Main questions:</p> <ul style="list-style-type: none"> • What is (and what caused) the problem? • How can we address the problem immediately? • How can we prevent the problem from recurring? • What course of action will we take and why? <p>Key advice figuring out root causes should be a key concern</p>
<p>Team-building meetings</p> <p>Intention: gaining a better understanding of one another and furthering cohesion among the group</p> <p>Main goals</p> <ul style="list-style-type: none"> • Improving trust among teammates • Helping the group feel valued and welcomed • Bolstering faith in the group's mission • Increasing dedication to getting the job done <p>Team-building exercises Games, talent showcases, group creative classes, team happy hours</p>	<p>Info-sharing meetings</p> <p>Intention: info-sharing meetings are a chance to educate the team in-depth on pertinent topics</p> <p>Main goals</p> <ul style="list-style-type: none"> • Use visual aids (pictures, video, etc.) • Engage the audience with questions • Weave personal stories into the information • Implement polls and surveys to capture attention • Use multiple speakers or guest speakers <p>Key advice Keep attendees focused</p>	<p>Innovation meetings</p> <p>Intention: thinking up designs, brainstorming content for an ad campaign and generating ideas about potential new products</p> <p>key advice</p> <ul style="list-style-type: none"> • The quantity of ideas beats quality at this stage • Refrain from overanalyzing ideas • Encourage "out there" thinking <p>Main goals</p> <ul style="list-style-type: none"> • Generated a ton of ideas for the project • Pushed the limits of the group's creativity • Gained excitement about the project that will spur further efforts

Important aspects about meeting

In the process of reading articles and people's insights about meeting, I highlighted those interesting aspects and put them together as the inspiration for the later stage.



Part of the interesting aspects

- Encourage movement
- Encourage equal participation
- Create an ambient connection to the space
- Knowing everybody's looking at you can cause social pressure
- Don't know the feedback from audience
- Have a clear and well-communicated agenda
- Over-Running and Over-Scheduling
- Design for thumbnail view
- ...

Gestures in remote meeting & physical presentation

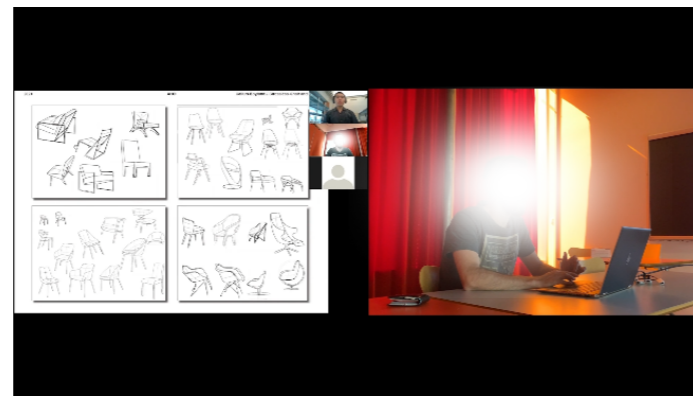
Because there is not much study about gesture differences between remote meeting and physical presentation, I conducted a workshop to have a brief view of the differences.

How to conduct the workshop

In the workshop, I asked the candidate to explain one of his slides to me both in the physical meeting room and through Zoom, the process was recorded so I could analyze it later. As a broad exploratory project, I haven't gotten any direction at this stage, so the workshop here is also different with those that requiring a lot of preparation and candidates, the purpose of the workshop here is having a glance of possibility behind the aspects in the short time.



Physical presentation



Remote meeting

Some interesting findings

- Beats gestures shows most in both physical and remote presentation
- Less iconic gestures in remote meeting
- Even user can point by cursor in remote meeting, he still use hand to point things subconsciously



Point with hands in remote meeting

Comparison and analysis

Because of my research about natural gestures, I can recognize those different type of gestures, I extracted those gestures from both videos and compare them. From there I can get some interesting findings.

designer's co-speech gestures

		Beats		Iconic			Metaphoric	Deictic	Cohesive	Emblem	Reflections
		Emphasize words/phrases	Uncover structure	Information on actions	Shape/ Size of an object	Spatial relation between two objects	Abstract ideas to visual	Pointing / direct attention	Continuation of a particular theme	can represent info without speech	
Describe a sketching (object)											<ul style="list-style-type: none"> • highlighting words is important and shows the most • how to show the rhythm of the speech in remote meeting? • are gestures can be categorized in natural type to speech gestures, it's even hard for people to distinguish, we also machine • the difference between pointing with cursor and pointing with hand? Besides image? the layer between human and object?

findings

problems

ideas

Review original theme

After those research, I can review my broad theme and narrow it down a bit, find a rough path to explore the possibility.

Unclear part

- Support or tool
- What kind of user? What kind of ideas?
- Remote meeting or physical presentation?

2 paths to explore

- Path 1 Support
- Path 2 Tool

Unclear part

My original theme is rough and broad, there are those parts that it's not clear.

Support or tool?

Explore how hand moving **gesture interaction** can support **user** expressing **ideas** among a group of audiences during the **remote meeting or physical presentation**.

What kind of user?

What kind of idea?

Remote or physical ?

Support or tool

“Support and tool” this concept may be quite abstract, the meaning of “support” is making use of human’s natural gestures to support explaining ideas, speaker can still perform those natural gestures unconsciously, but there will be some effects happening to strengthen their expression. In the contrast, “tool” means defining some gestures and user can use them deliberately as tools, those tools will trigger some effects to support their talking.



Support

Making use of people’s natural gestures in communication and support people expressing ideas

Advantage:

- Don’t need to learn new gestures
- Won’t influence the original using of co-speech gestures

Disadvantage:

- The same gesture can have multiple manifestations, it can be hard to detect
- Using camera and ML means user’s hands have to show up



Tool

Define new gestures and use it as a tool to support people expressing ideas

Advantage:

- Flexible, there will be more possibility according to different topic

Disadvantage:

- User needs to learn new gesture
- May conflict with the natural gestures in the conversation

What kind of user? What kind of ideas?

Topic/idea rather than user

As there are lots of different people joining meeting, in the beginning of the project, I was struggled to define a user group. After the reflection of the theme, I think it’s not about user, it’s about topic/ideas. Because the same users may talk about different topic. For example, two designers can talk about design in the meeting, they can also talk about music or history. Instead of asking what kind of user, I would start exploration by what kind of topic/idea.

Observation & ideation

When it is determined that topic is the key, I also did a simple ideation phase about what kinds of topics we can have in the meeting, then I also conducted the observation of people talking about different topic.

Remote meeting or physical presentation?

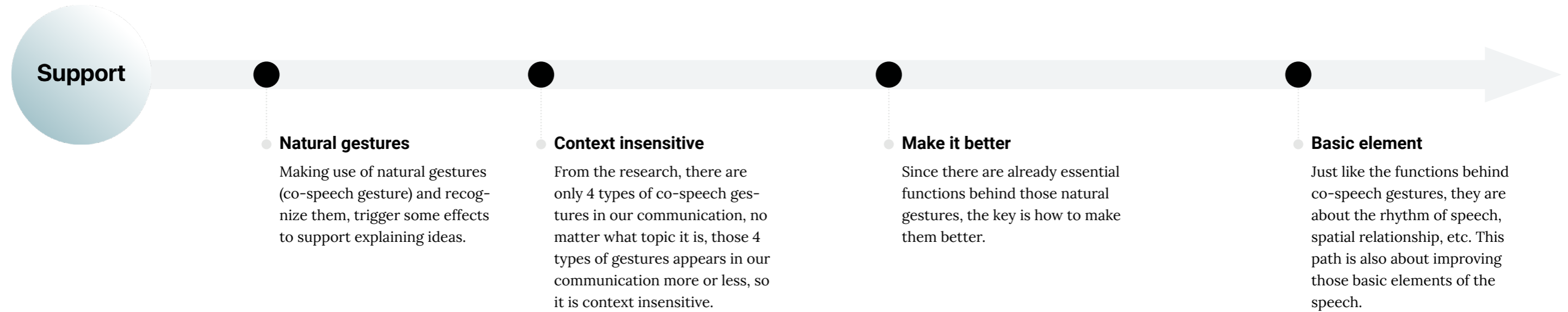
Remote meeting

Since it’s hard to explore from both aspects, I chose to start with remote meeting, because the gesture recognition technology I want to use is based on camera and there is already a camera as the input device in remote meeting.

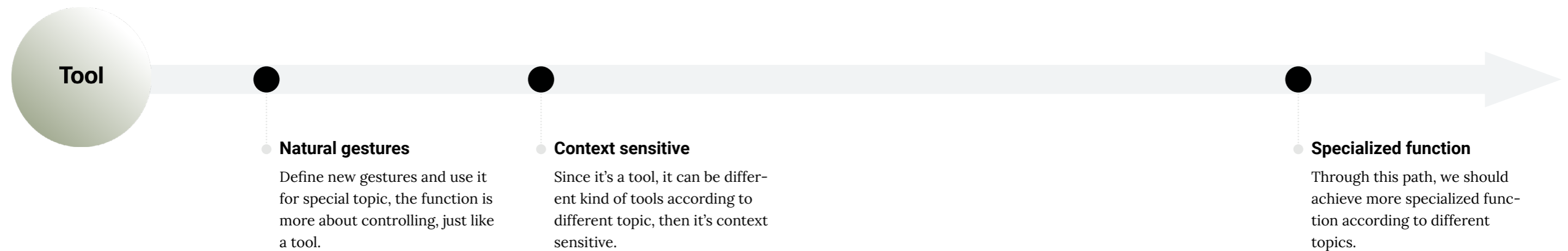
2 paths to explore

After those reflections of my original theme, I came up with these 2 paths to explore the project.

Path 1: Support



Path 2: Tool



Ideation 1.0

Along with those 2 paths, I started the first ideation phase. In this phase I came up with 8 rough sketches, some of them are video sketches (edited video) and some of them are programmed working prototype, in this way I can sketch quickly and also have a deeper understanding of what this technology can do.

Ideation 1.0

- Make use of beats gesture
- Pointing with hands & character
- Drag words from mouse
- Describe color?
- Music, create beats?
- Spatial relationship
- Make a list
- Create basic shapes
- Insights

Watch video then read text

Because gesture is dynamic, it's hard to explain by the static image or words, please click the hyperlink under the image or scan the QR code to watch the video. Watching the video first then viewing the text description, it will make more sense.

If clicking every time is too troublesome, this hyperlink contain all my sketches in both ideation 1.0 and ideation 2.0

<https://allsketches.netlify.app/>

Ideation 1.0

Make use of beats gesture

Path 1: Support / beats



Please scan the QR code or click the hyperlink under the image to see the real effects



<https://ahocloud.box.com/s/hswntf9ftcw5mpzxyjsji4pqgcr5pt>

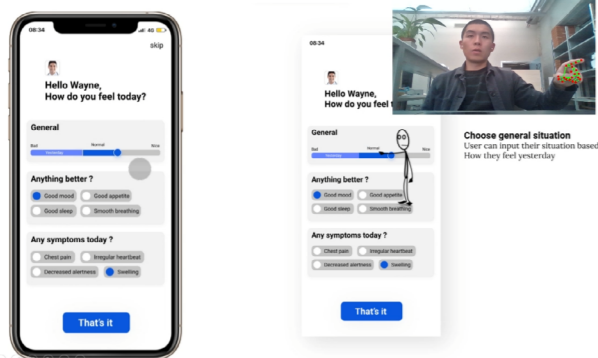
Making use of beats gesture, every time people have obvious hands up and down movements, that means the speaker is saying the keywords.

- Recording the speech and put that keywords on the screen.
- Or generate a list of keywords as notes after the conversation.
- According to hands movements, use this wave effects to visualize the rhythm of the speech.

This is a working prototype.

Pointing with hands & character

Path 1: Support / deictic



<https://ahocloud.box.com/s/61eo484dldx610stepd28xbdno70cc61>

From the research phase I found that even in remote meeting, people tend to point with their hands, maybe we can just point with our hands. Moreover, instead of using the cursor, perhaps using a character to point, through this way, we can simulate the physical presentation.

This is a working prototype.



Similar effects have already been used by some YouTubers, the difference is those keywords are edited after the video.

Observation



It seems that pointing in physical presentation draws more attention than pointing with cursor in remote meeting, maybe in the physical presentation, there is a hierarchical relationship between people and screens which may also help to draw attention.

Drag words from mouse

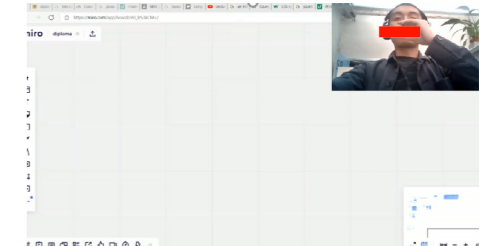
Path 2: Tool / context of ideation



<https://ahocloud.box.com/s/5b3t1m7qfxdaxhfxia42azx9hkv9tyfu>

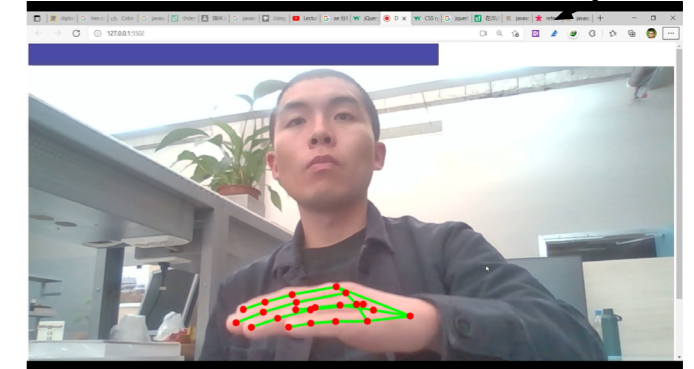
When the topic is about ideation, maybe speaker can perform certain gesture near their mouth and drag the words he said, then put it on the screen. Show that gesture near the mouth, the machine will start to recode sounds and transfer it to screen.

Say color comes out color box



Describe color?

Path 2: Tool / topic of color

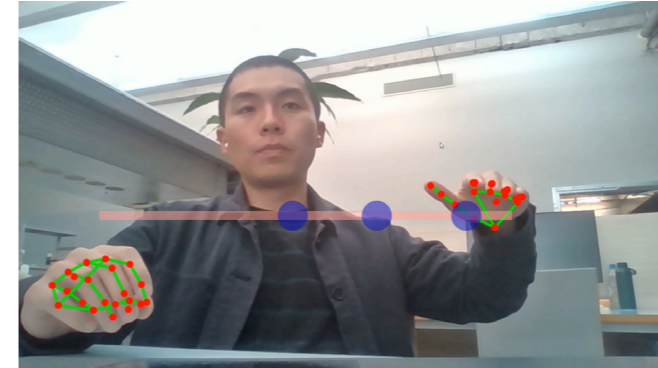


<https://ahocloud.box.com/s/e1aa6txphkymbdzjew0cjuw20ejm6mpa>

Go further with dragging words from mouth, when user says "color", instead of coming out the word "color", there will be a color box, user can use hands to control the color of the color box. This is a working prototype.

Music, create beats?

Path 2: Tool / topic of music



<https://ahocloud.box.com/s/nfnta3exvrww06kyb5nzjefr31in0cbw>

Go further with dragging words from mouth, when user says "beats", a beats bar will come out and user can create beats with his hands. It can be useful when talking about music.

Observation

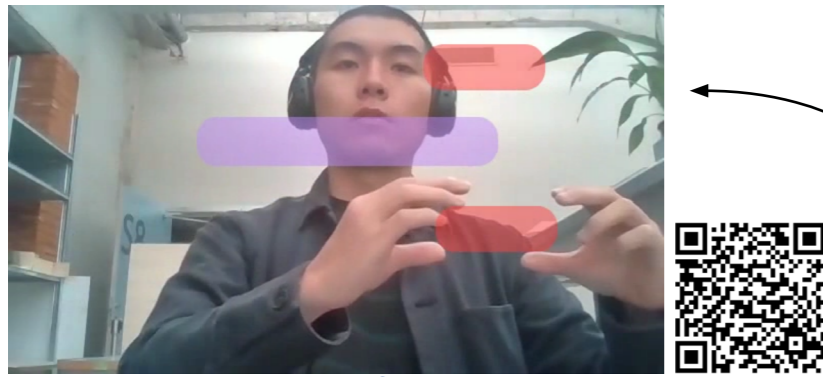


In the music class, the teacher uses hands to help create and explain beats.

Ideation 1.0

Spatial relationship

Path 1: Support / iconic

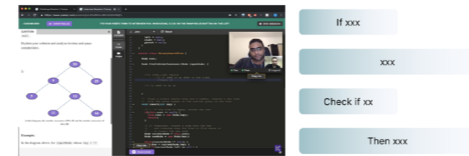


<https://ahocloud.box.com/s/7xh44699fl2jhhpcsgl5e3y6iqk5lgkn>

When speaker's hands get close, there will be a shape created between the hands, with those shapes, the expression of the spatial structure would be strengthened.

Interestingly, I started this idea from tool, just under the coding topic, but it leads me to the support path.

Observation



The observation of a coding interview. In that interview, they didn't talk about exact coding piece but the logic behind, first check something, then do something, there is an obvious spatial relationship behind it, and from the research, we do have iconic gesture to explaining spatial relationship.

Create basic shapes

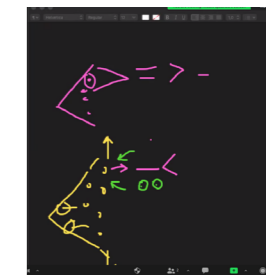
Path 2: Tool / context of sketching



<https://ahocloud.box.com/s/zad8ik2f0rvklgiqp0jgt2bczb5faav>

Maybe the user can use gesture to create some basic shapes and explain ideas. It doesn't need to be complicated shapes, from the observation, we just need some basic shapes.

Observation



The observation of people explaining ideas with sketching, here is the sketch that my supervisor did to explain double diamond model to me during the remote meeting.

Make a list

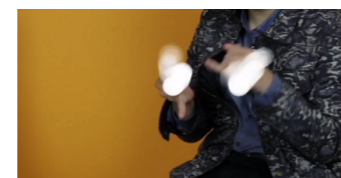
Path 1: Support / beats



<https://ahocloud.box.com/s/t4h0rj36w58dr1hg905wfw4rznl0xwdk>

Every time the user uses this gesture, the machine will record what the speaker says and put it on the screen, so the user can make a list on the screen easily.

Observation



A common gesture people use when they say a list of things, for example "I ate banana, apple, hamburger this morning." Maybe not exactly the same as the picture shows, but this gesture is still very common. From the research I did, it should be the beats gesture.

Insights

After the ideation and testing of Ideation 1.0, I found a paradox, I wanted to detect people's natural gestures in the beginning, but once people find that their gestures can make some effects, next time they will do it deliberately, then it's not natural gesture anymore. So I had to dilute this "unconscious natural gesture" idea, the meaning of natural gestures later are not those co-speech gestures, they are the gestures people would perform without too much thinking.

Ideation 2.0

In order to explore the full possibilities under the theme, I didn't stop at the Ideation 1.0 phase, in fact after those sketches, I conducted a workshop to help generate more ideas.

The sketches in this phase are consist of edited videos, prototypes and hand sketches. Some of them will be explained in detail with videos, the rest of them I will just roughly go through them with description.

In fact, every idea is accompanied by a video, if you want to watch all of them, please click this link

<https://allsketches.netlify.app/>

Workshop

More ideas

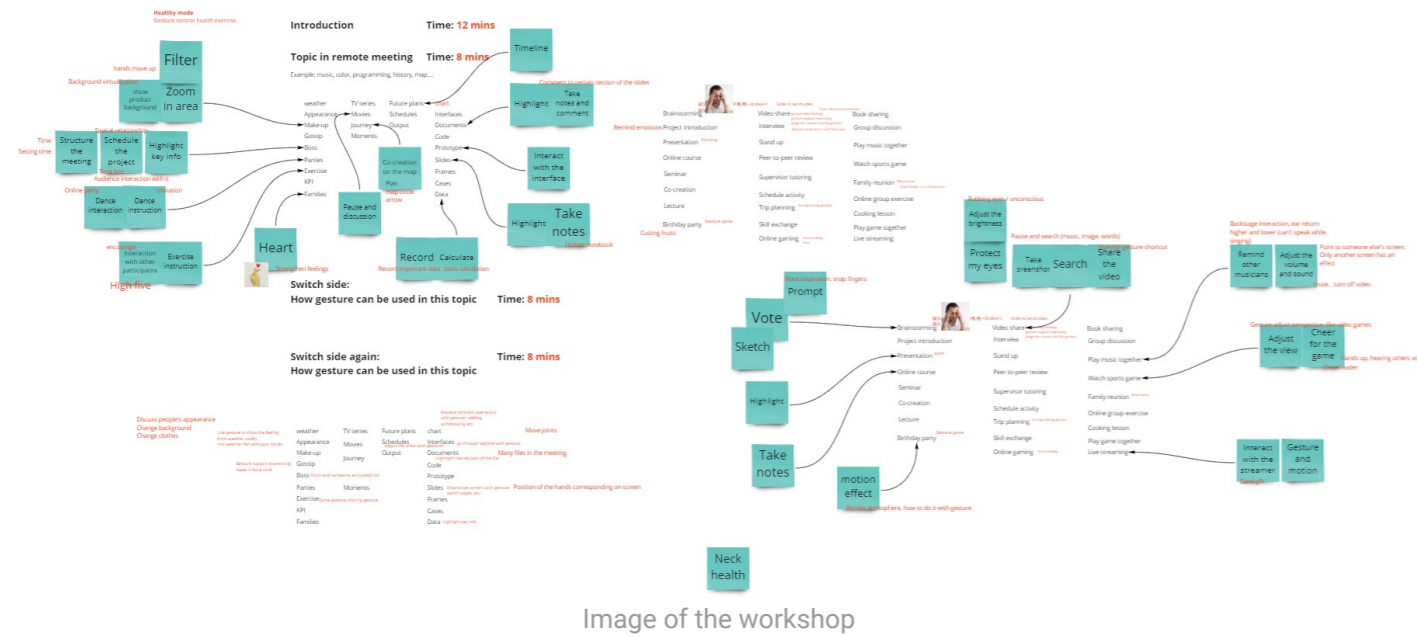
- Call out notes
- Set a timer
- Degree
- **High five with gesture!**
- **Whisper/Gossip**
- Gesture calculation
- Idea about general problem
- Other ideas in sketching

Watch video then read text

Because gesture is dynamic, it's hard to explain by the static image or words, please click the hyperlink under the image or scan the QR code to watch the video. Watching the video first then viewing the text description, it will make more sense.

Workshop

To generate more ideas, I conducted a workshop, in that workshop, candidates were asked to first think about what kind of topics we can have in the meeting, then switch side and view other's topic, think about what gestures can do about it.



More ideas

Call out notes

Path 1: Tool / context of presentation



During presentation or explaining ideas, user can prepare some notes or inspiration images in advance, then call it out by snapping the fingers. Those notes will be hidden to other audiences.

Set a timer

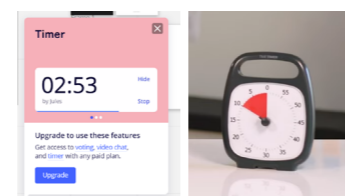
Path 2: Tool / context of workshop



<https://ahocloud.box.com/s/fkb4re1kw2vrqrckhn2065y7ic2qs4d7>

Maybe in the remote workshop, we can set a timer with this gesture.

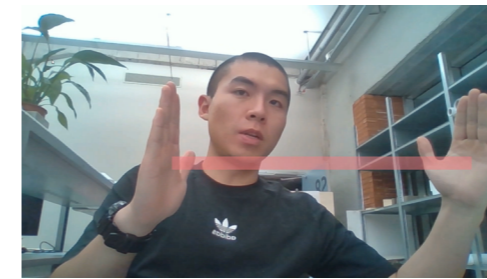
Observation



It is common to set a timer in both remote and physical workshop.

Degree

Path 1: Support / metaphoric



Use gesture to transfer abstract degree into the length and color, for example, "it's extremely great" then the bar will be long and red, "it's slightly different" then the bar will be short and gray.

High five with gesture!

Path 2: Tool / atmosphere



<https://ahocloud.box.com/s/4zsjhbf73e6ku3hn6c7yzndrh8sv0v2t>

From the research I did about meeting, atmosphere is important, normally we can thumb up towards other people, maybe we can also high five with other audiences. Waving hands to activate high five function and high five with the others as the video shows. Compared with thumb up this one-way expression of emotion, high five can promote the interaction between audiences.



Please scan the QR code or click the hyperlink under the image to see the real effects

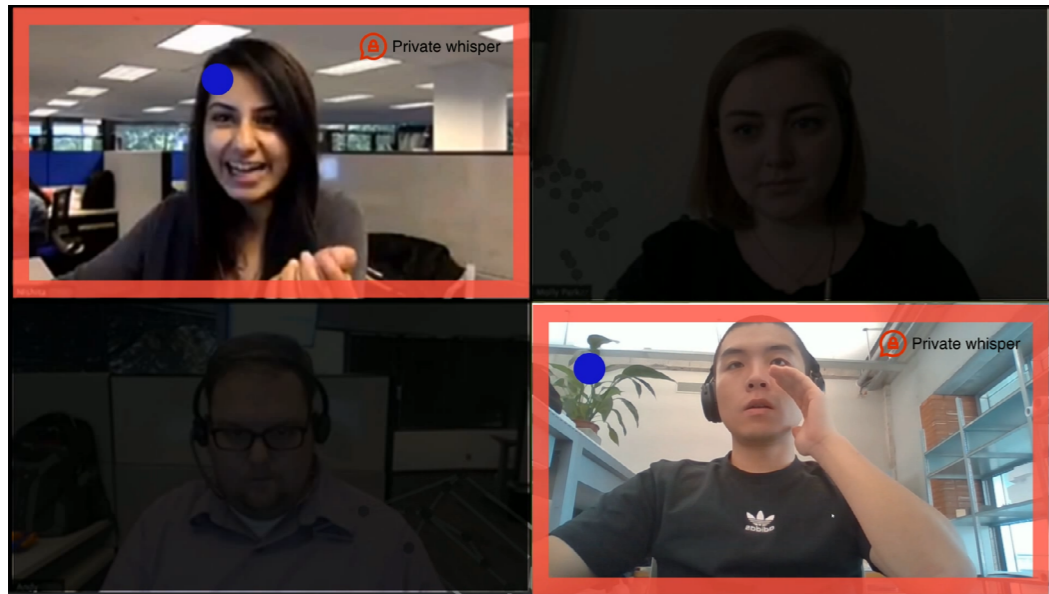
Inspiration



In the video game DOTA2, players can high five with their teammates or enemies, it's a successful function that creates a cheerful atmosphere.

Whisper / Gossip

Path 2: Tool / Under the context of whisper



Please scan the QR code or click the hyperlink under the image to see the real effects

<https://ahocloud.box.com/s/e456i1tzlzwkf8ctk9fpakvlhrp4q44x>

From the research of meeting, gossip is important in the working place and that inspires this ideangossip/whisper. As the video shows, the user can use the certain gesture to trigger the whisper function, then use another hand to choose the target, after that, the conversation between the user and the targets will become private conversation. From the other audience's view, the user and the target's camera video will be blurred.

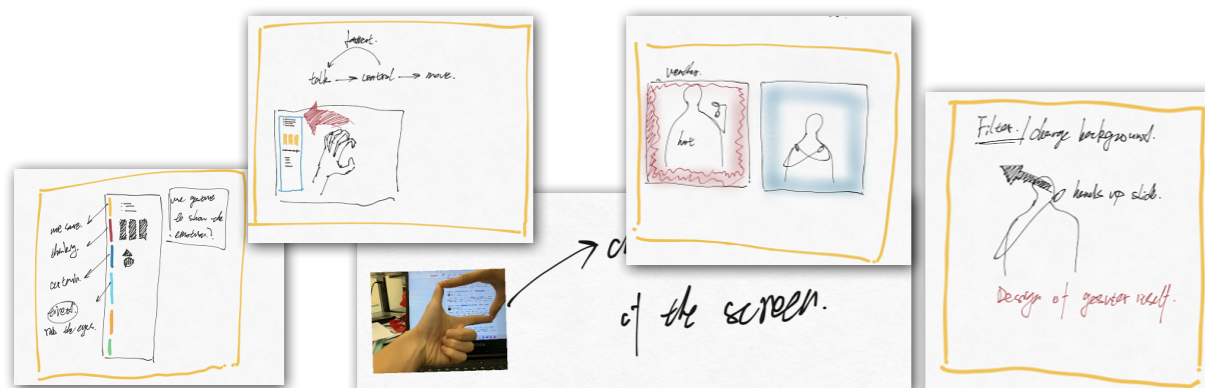
Inspiration

Even when the content of gossip is dull and disconnected from your day-to-day obligations, gossiping remains a source of bonding, which everyone craves at a biological level. In fact, gossip literally soothed people in one University of Toronto study, and was found to encourage cooperation. Through gossip, we all absorb cultural rules and warnings: Because it was spread through the grapevine, you are spared a first-hand experience of Danny in accounting's underhandedness, or Maria in marketing's short fuse.



From the research of meeting, gossip is important in the working place and that inspires this ideangossip/whisper. Also in the classroom, students whisper with each other a lot. "What's he talking about?" With the gesture shows in the image.

Other ideas in sketching



Gesture calculation

Path 2: Tool / topic of data



When taking about data, maybe we can use hands to do some quick calculation.

Idea about general problem

Here are two models I came up with to solve the problem, there are also working prototype.

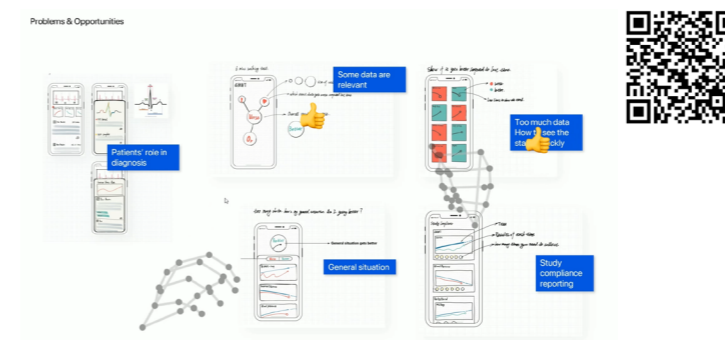
Model A - 3 layers



<https://ahocloud.box.com/s/o8qys0nwg709tzb6asd6ra8zl2sv218>

When sharing slides, there will be three layers, at the bottom is video layer, which shows the camera video of the presenter himself, in the middle is slides layer, it will be slightly transparent, so the presenter can see where his hands are, but the slides will still be solid for other audiences. On the top is the effects layer, which is the place where shows the effects the presenter makes with gestures.

Model B - 2 layers



<https://ahocloud.box.com/s/4ma4odlxgu1pik8l656hwzd4cckukvdi>

In this model, there will be 2 layers: effects & hands layer, slides layer. The presenter can only see their virtual hands on the screen, under that is his slides.

Testing & discussion 1.0

During the ideation phase, I also discussed those ideas and tested prototypes with other people and the experts in remote meeting this field.

Testing & discussion 1.0

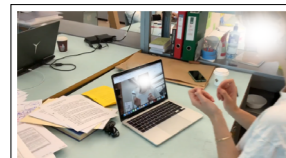
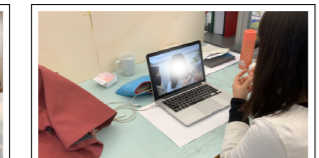

- Test with regular people
- Test and discuss with experts
- Test and discuss with supervisors

Test & discuss 1.0

Test with regular people

Showing video ideas with other people and testing the prototype with them. Then write down and organize the feedback.




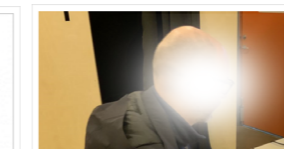
 <p>Eirunn Student (interaction design)</p> <p>Feedback</p> <p>Making list</p> <ul style="list-style-type: none"> • Don't have similar experience before, hard to compare • Useful • Other audience can interact with it • Don't know when to use it before explaining • Won't use it for too many times <p>Spatial relationship</p> <ul style="list-style-type: none"> • Don't know when to use it before explaining • Won't use it for too many times 	 <p>Chia Han Student (service design)</p> <p>Feedback</p> <p>Making list</p> <ul style="list-style-type: none"> • Don't know how use it in the beginning • What will happen afterwards, further actions • Can trigger other things such as "video, picture, etc" • If its a conversation and all the people can talk, what will happen when other people talk • Useful in class <p>Spatial relationship</p> <ul style="list-style-type: none"> • Don't know when to use it • Compare, size and shape 	 <p>Callum Student (industrial design)</p> <p>Feedback</p> <p>Making list</p> <ul style="list-style-type: none"> • Explain how to use (touch fingertips rather than joints) • Special gesture may hide the palm and cause mis-triggered
---	---	--

Some interesting feedback and reflections

- **How to explain the function to users**
During the testing, I find it hard to explain the function to the user, especially describing the gesture.
- **How can the user know the purpose of the function**
There is a prototype that people can create a shape when hands get close, I made this function to help user explaining spatial relationship, but it's hard for user to know it if I don't explain. In fact, people use it in quite different way, they use it to highlight things, which may not be a bad thing.
- **Further operations after effects**
After making some effects on the screen, what other things the user can do about it.
- **Talk to machine or audience.**
In one of the remote discussion, I showed the prototypes to the candidate, he told me "I feel like you are talking to the machine rather than me" that's an interesting point.

Test and discuss with experts.

I showed the ideas and tested the prototypes with the experts in meeting this field. I talked with Mona and Therese, UX designers from Huddly (company that producing cameras for meeting), I also discussed with Fredrik, creative director from Bekk, who has a great knowledge about meeting this field.

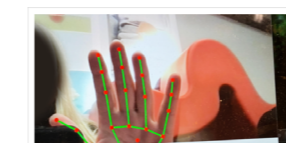
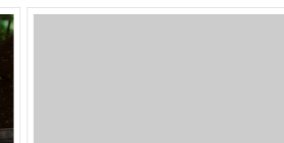
 <p>Huddly Produce remote meeting camera Mona Kleven Lauritzen</p> <p>Feedback</p> <p>Making list</p> <ul style="list-style-type: none"> • Context rather than summarize persona • Ask user to pick which function is useful • Think about roles in the group (someone speak a lot) • Presentation without slides • Give a nudge to someone • Position of the camera • Use it in physical presentation 	 <p>Fredrik Matheson Creative Director - Bekk</p> <p>Feedback</p> <p>General</p> <ul style="list-style-type: none"> • Gesture and audio • Chapter of the conversation, change scenery • Role in the meeting • Memorable of the gesture • How to show the status of the system (e.g. if the gesture function is activated?) • Presentation • Management of the meeting • Feedback of audience • mmhmm.app
--	--

Some interesting feedback and reflections

- **Think about the role in the meeting**
There are different roles in the meeting, speaker, audience, leader, etc. What will they do with gestures?
- **Position of the camera**
The technology is based on camera, the position of the camera would have a huge influence towards those functions. However, I think this can also be an opportunity, having some mobile camera can make those ideas also working in physical presentation.
- **Is those function/gesture memorable?**
Some gestures may be natural to use, but when we ask the user to remember that gesture and perform it deliberately, it can be hard to remember, because we are too used to it.

Test and discuss with my supervisors

Some interesting feedback and reflections

 <p>Mosse Teacher (Design)</p> <p>Feedback</p> <p>Making list</p> <ul style="list-style-type: none"> • Where to notify user that the function has been triggered. • Other audience can also interact with it 	 <p>Lars Teacher (Design)</p> <p>Feedback</p> <p>General</p> <ul style="list-style-type: none"> • How user can understand the function • What the audience can do • How the user can understand the purpose of it • Ask user if they can understand and how they will use it
---	---

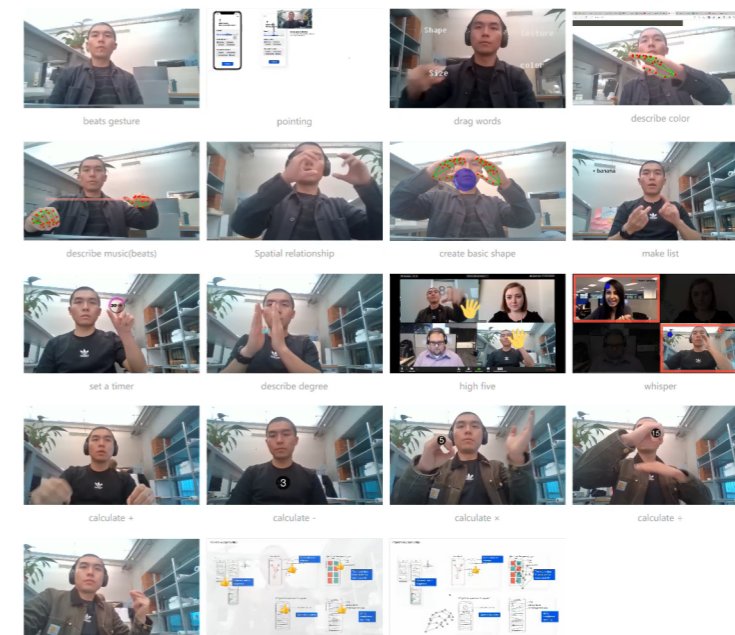
- **How to show the status of the system (triggered or not)**
In the prototype of making lists, every time the user uses that gesture to trigger the voice recognition, there will be an orange dot on the corner of the screen to indicate that the function has been triggered, however, during the testing, I found that people's attention is on the hands rather than some corner of the screen.
- **How can audience interact with it**
Once the speaker created some effects on the screen, how can the audience interact with it?

Evaluation

In the ideation stage, lots of ideas are generated, they are relevant to various aspects of the meeting, just like different functional pieces. However, I cannot explore all aspects but choose some of them to go deeper. So the problem is how to evaluate those ideas?

Evaluation

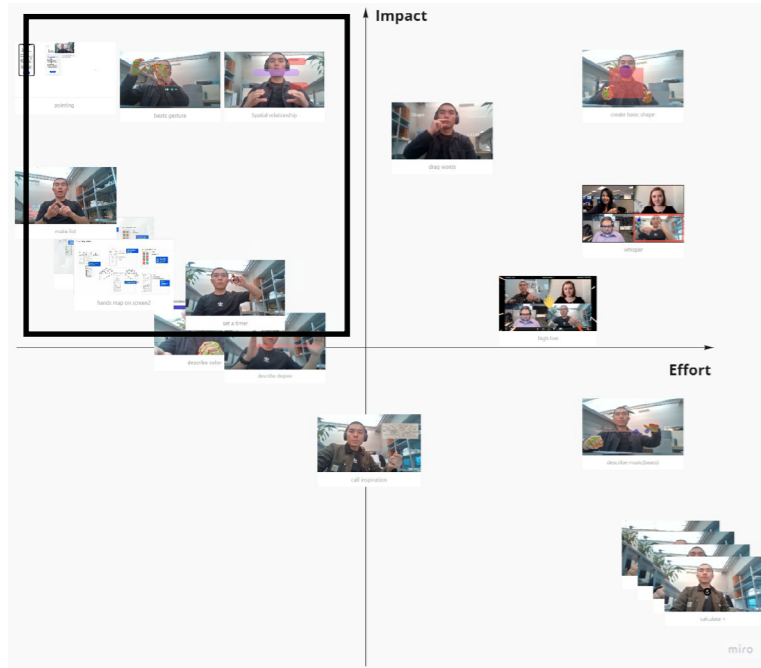
- Impact / effort matrix
- Prototyping
- Take apart the ideas
- 5 stages
- Custom evaluation matrix
- Map out ideas & go deeper
- Foundation



Evaluation

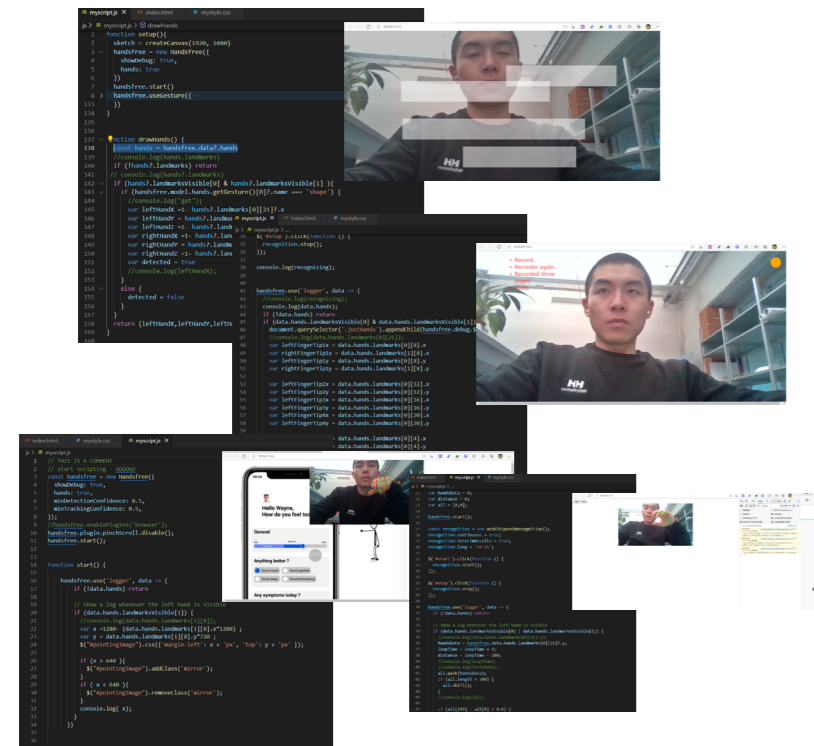
Impact / effort matrix

To evaluate those ideas, I first put them into this impact/effort matrix, try to figure out which idea has bigger impact and require fewer efforts. Obviously, the top left area is the sweet box, the ideas in this area are what I should turn into more refined prototypes.



Prototyping

It's hard to imagine or experience gesture interaction without really using it. Furthermore, if not making the working prototype, it's also hard to know what kind of problem there will be.



Take apart the ideas

I also tried to put all ideas together and take it apart to see

- What gesture triggers the function?
- What gesture confirms the effects
- Will them interfere each other?

Through this way, I want to find some pattern behind those ideas.

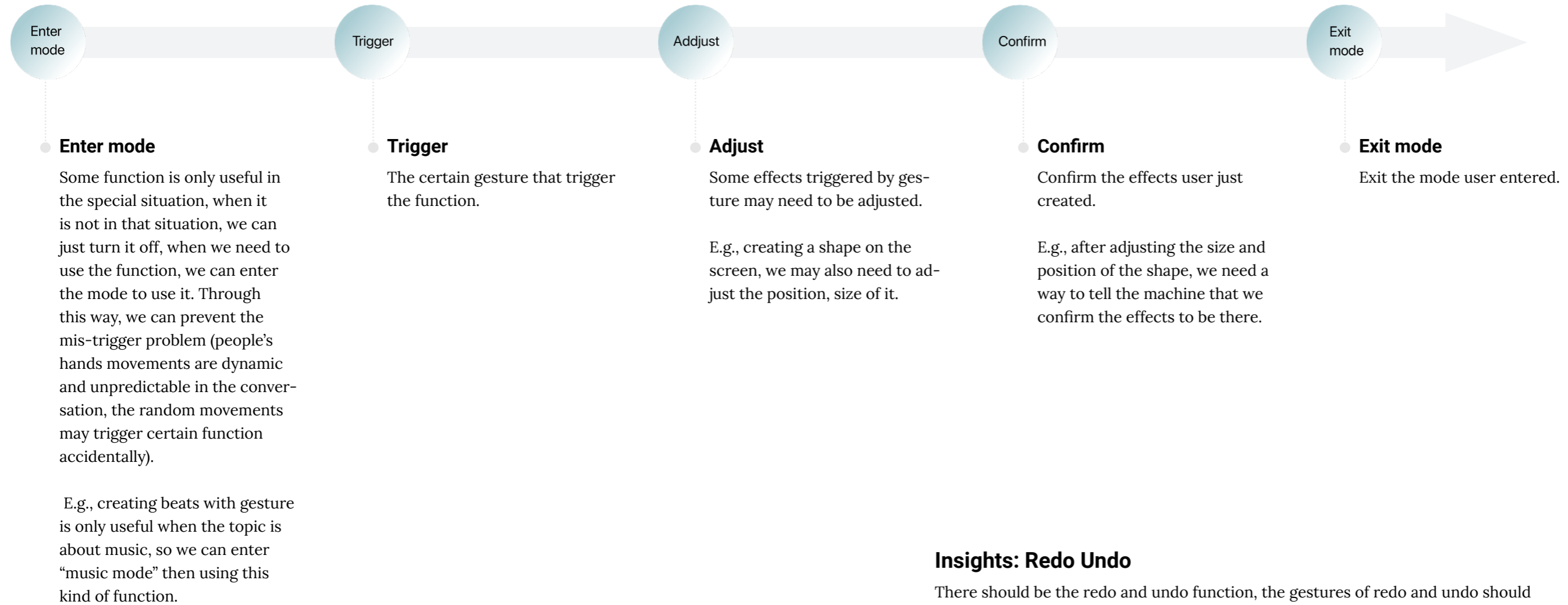
	Start function	Trigger effects	Adjust	Confirm	End function	Undo	Redo	Input source	How many hands
		Up & down of hands						Gesture & language same time	1/2
	Not sure	Hold (not touch)	Distance between palms	Z index	Not sure			Gesture	2
		Finger tips touch together	Adjust position of effects	Finger tips leave				Gesture & language same time	2
	not sure	Hold (not touch) (straight)	Distance between palms	cancel the gesture	Not sure			Gesture	2
	not sure	Hold (shape)	Distance between palms	Z index	Not sure			Gesture	2
			Drag Up & down of hands	Z index (push?)				Gesture & language before language	1
			drag	End recording (time)				Gesture & language before language	1
			Drag hands up and down	Z index (push?)				Gesture and language before language	1
			Position of fingertips					Gesture	1
				Z index (push)				Gesture	1
	Workshop mode?		Drag rotate	Z index (push)				Gesture	1
	Presentation mode?		Drag					Gesture	1
			Drag 2 number get close					Gesture and language before language	2
			Drag 2 number get away					Gesture and language before language	2
			Drag 2 number draw angle					Gesture and language before language	2
			Drag 2 number up and down					Gesture and language before language	2
			Position of fingertips					Gesture	2

Evaluation

5 stages

Because of programming them out and taking apart those ideas, I do find a pattern behind them, according to that, I came up with this 5 stages model.

When designing this kind of gesture triggered function, we can think about it from these 5 stages: enter mode, trigger, adjust, confirm and exit mode.

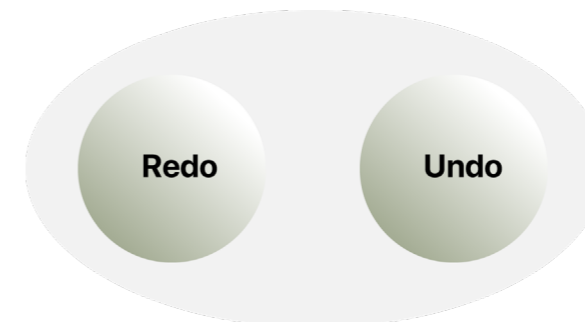


Depends on the situation

Although we need to think about these 5 stages, but that doesn't mean we always need all 5 stages. E.g., making a list with gesture, we don't need to confirm it. Some function can also be turned on all the time, then it does not require entering mode.

Insights: Redo Undo

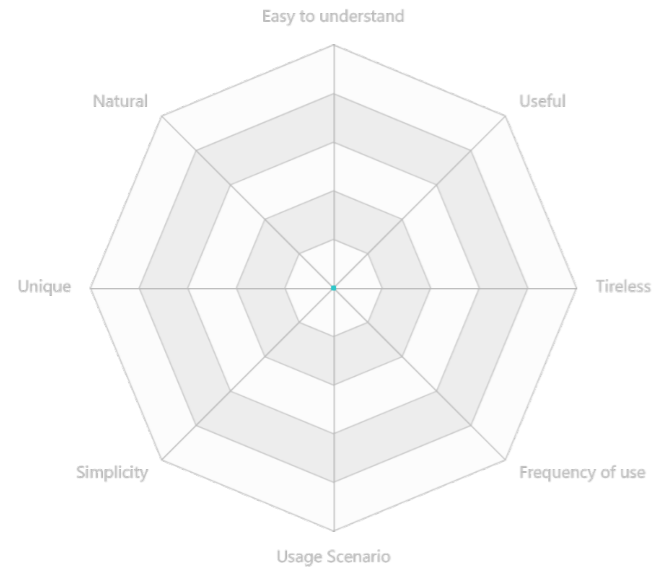
There should be the redo and undo function, the gestures of redo and undo should be same no matter what mode the user is in.



Evaluation

Custom evaluation matrix

Because of prototyping and taking apart those ideas, I gradually found out which aspects about the idea we should pay attention to. Based on that, I designed a custom evaluation matrix, so I can evaluate those ideas in a better way, there are 8 dimensions in this matrix.



Understandable

How easy for user to understand the purpose of the function. Moreover, how easy for the other audiences to understand the effects.

Natural

How similar is the trigger gesture to our natural gesture (natural gesture is a vague concept here, as mentioned before, by natural gesture I mean those gestures that people use without too much thinking.)

Unique

How unique the trigger gesture is, the more unique, the harder that function will be mis-triggered.

Simplicity

As the 5 stages model I mentioned, how many stages does the function require, the less, the better.

Usage scenario

How many scenarios can the function be used? Some functions can only be used under the special topic while some can be used in general talking.

Frequency of use

In the single scenario, how many times will the function be used. E.g., making a list function may only be used in the beginning of the meeting.

Tireless

How tired is it to perform the trigger gesture. Although It's relevant to the frequency of use, it also relevant to how high we need raise our arms etc.

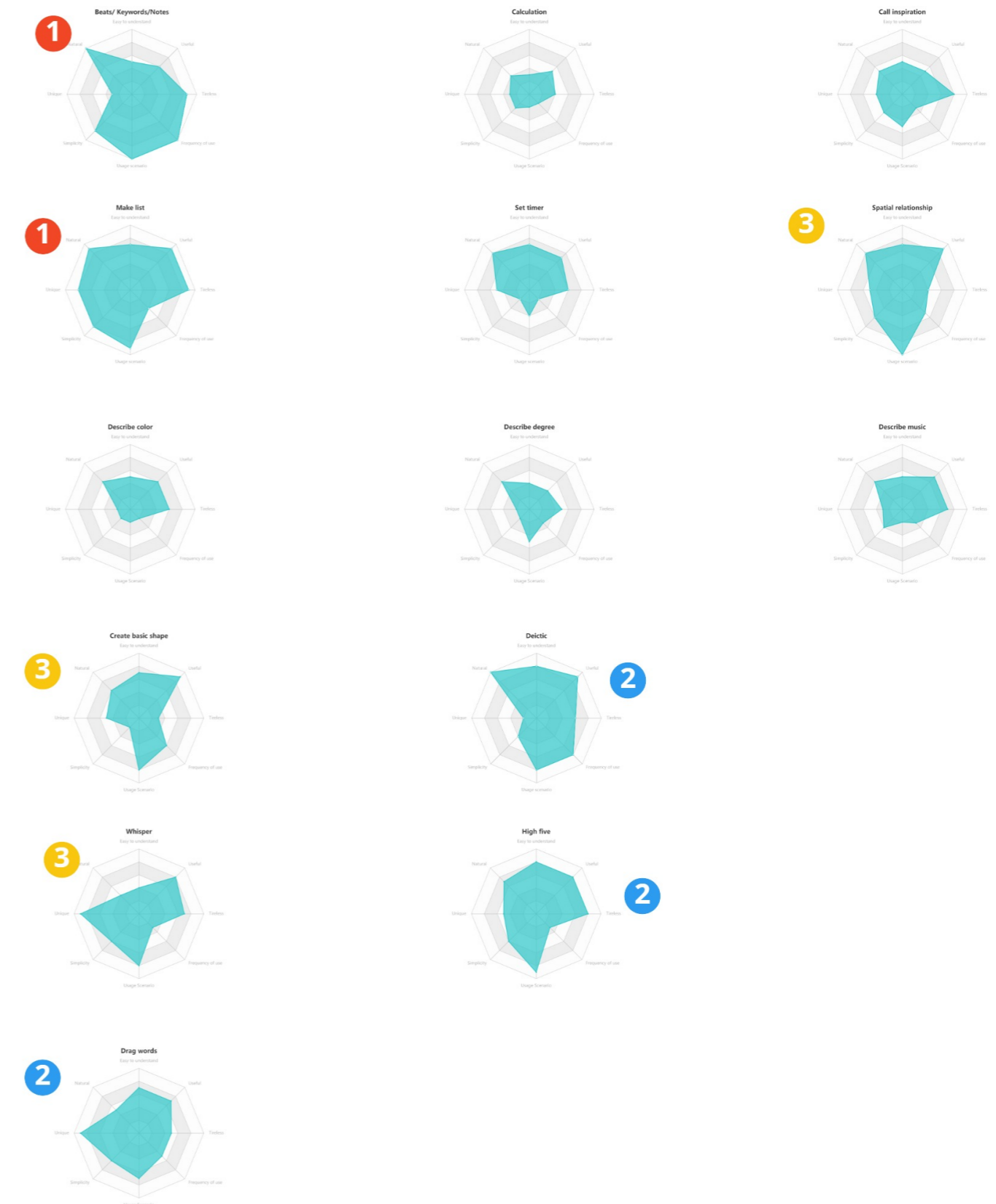
Useful

Overall, how useful is this function.

Evaluate with dimensions

Put those ideas into this custom evaluation matrix and evaluate them with the eight dimensions. Highlight those ideas with larger area.

Generally, the idea with larger area should be better, but that doesn't mean those functions with smaller area are bad, those functions may be extremely useful under some special topic.



Evaluation

Map out ideas

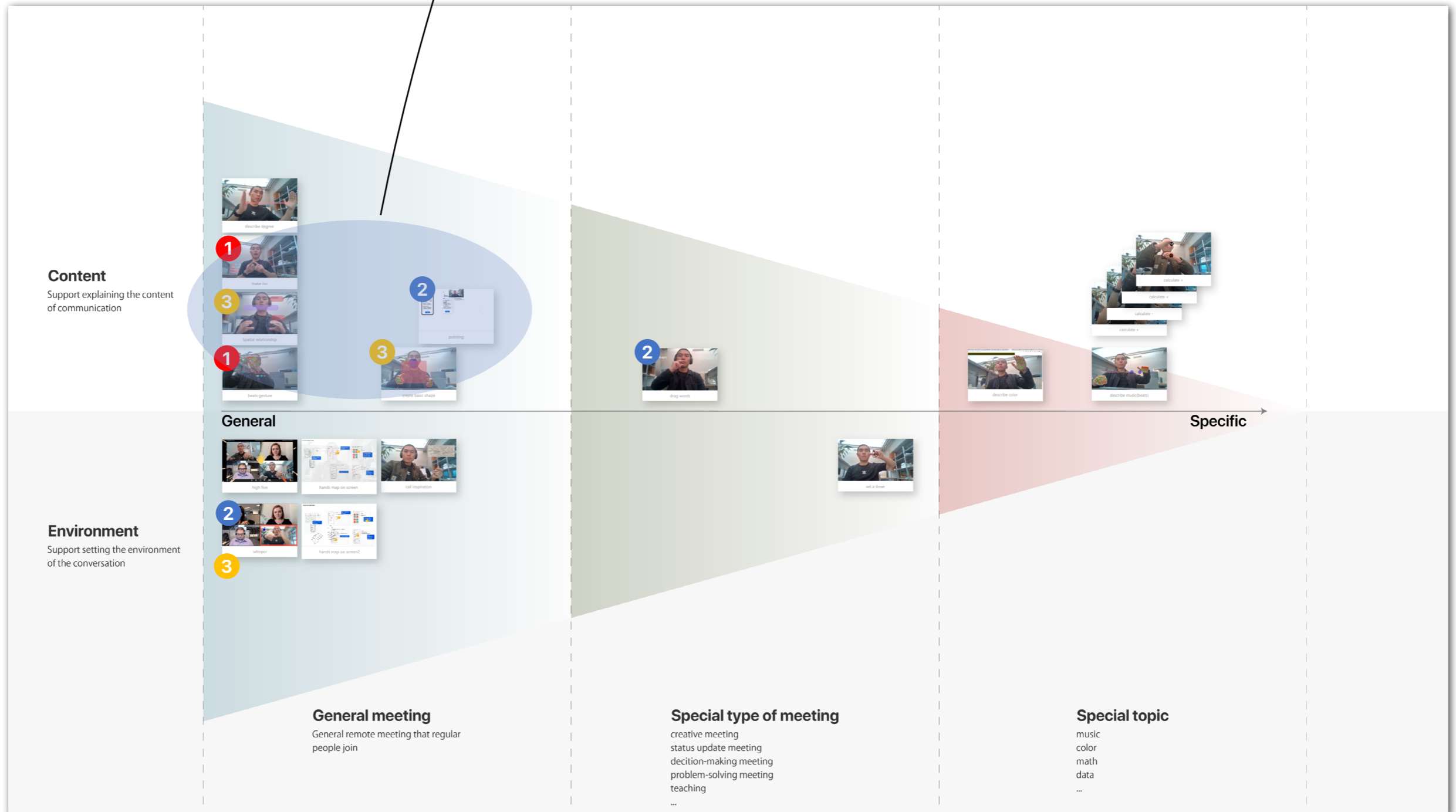
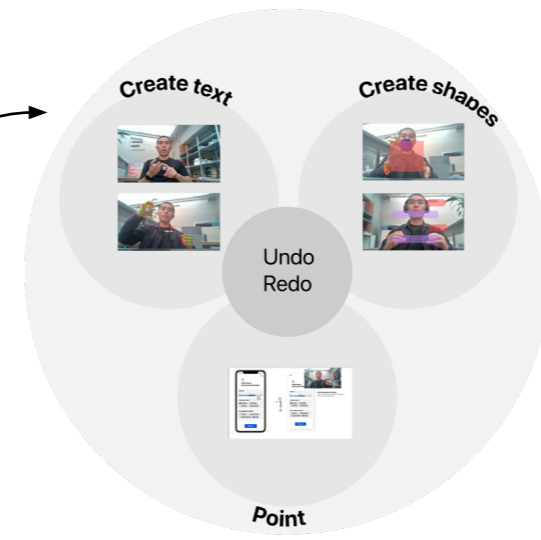
To have a more comprehensive view of all my ideas, I mapped out all ideas. From the map we can have a clear vision that some ideas are for general meeting, some are for special type of meeting, some are for the special topic. Ideas in the upper part are designed for supporting explaining the content, for example making a list. Ideas in the lower part are designed for supporting setting the environment of communication, for example whisper function and set timer function.

Results from custom evaluation matrix

Put the results from custom evaluation matrix into this map, we can find that the orange area has the highest scores. And that's the place I choose to go deeper.

Go deeper

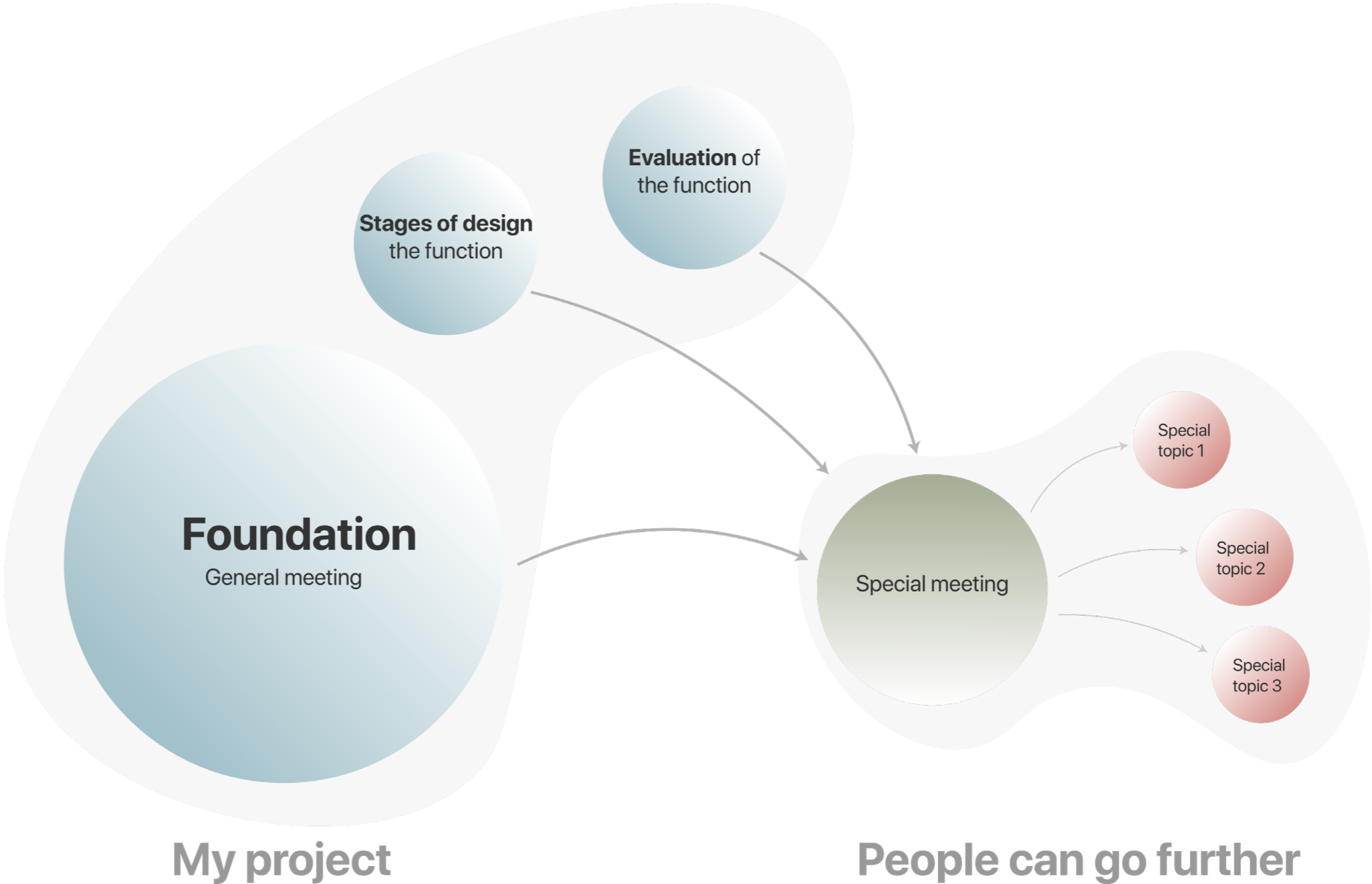
This area is the place I will go deeper, to be more specific, it's gesture interaction support regular user explaining ideas in general meeting. The ideas in this area can be categorized into three groups: creating text, creating shapes, pointing.



Evaluation

Not just combine, but deeper exploration

Even there are already some sketches in that group, but I won't just simply combine them together, but go deeper, explore how the audience can interact with the effects, what will happen if the speaker is changed, how it will be like in those remote meeting platforms which speaker sees other audiences' face rather than himself.



Foundation

Within that chosen area (supporting regular user in general meeting), an important aspect of my project is building a foundation during the exploration of using gesture recognition in the general meeting, finding the potential problems and exploring potential solutions, along with the 5 stages method of designing this kind of function and the evaluation matrix, people can go further and explore their more special functions.

Ideation 3.0

As an exploratory project based on the new technology (recognize gestures through camera with machine learning), I started this Ideation 3.0 by combining sketching, programming and testing.

Ideation 3.0

- Sketching with programming and testing
- Hand sketching
- Programming and testing
- Extra research
- Principle

Sketching with programming and testing

Sketching

Drawing out those ideas on the paper.

Programming

Although it's called "gesture recognition," in fact it only gets coordinates of our hand joints, there is a huge gap between getting coordinates of fingertips and really using gesture to interact with the machine. Without programming it out, we don't know what is possible and what problems we may meet. Moreover, sketching how can we use gestures is totally different with really experiencing them.

Testing

The testing part in this ideation phase is different with usability test, the purpose is not deciding which function is perfectly usable now, it is for experiencing that possibility, inspiring more ideas and finding the potential problems.

Extra research

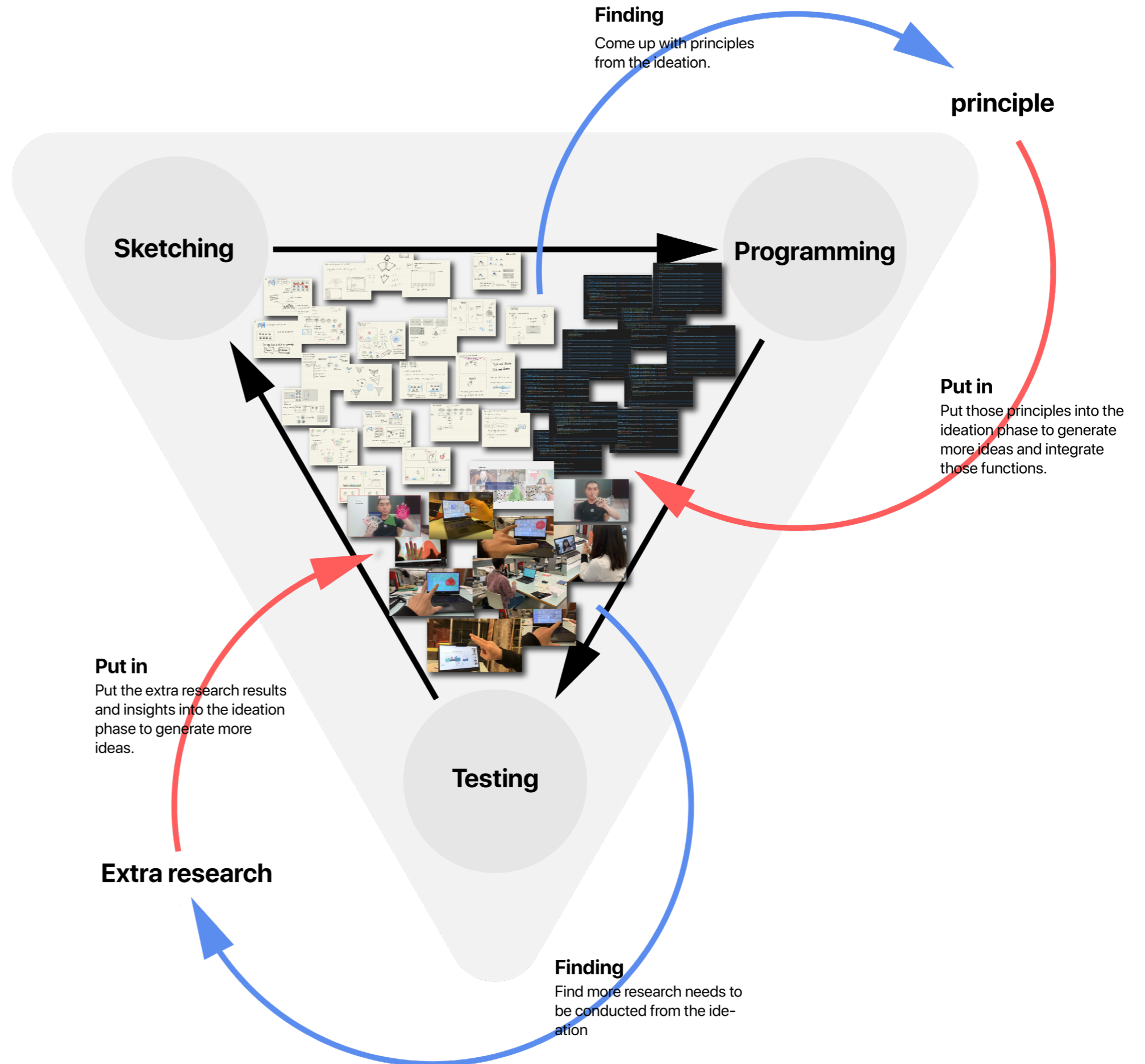
Sketching, programming and testing, those three parts are connected together, through the back and forth between them, I came up with many ideas from both big view and the detailed view, because of that, I found that there was more research needs to be conducted.

Principle

Except generating ideas, from the ideation phase, I also came up with some of my principles based on my insights, along with those principles, more ideas are inspired, because of those principles, I can integrate those ideas.

Nonlinear process

The exploration process here is nonlinear because there are no problems or goals in the project, before the exploration, I don't know what exact function there will be, what's possible, what research I need, but the purpose of this project is also not giving the only answer to those questions, it's more about showing the possibility, building a start point.



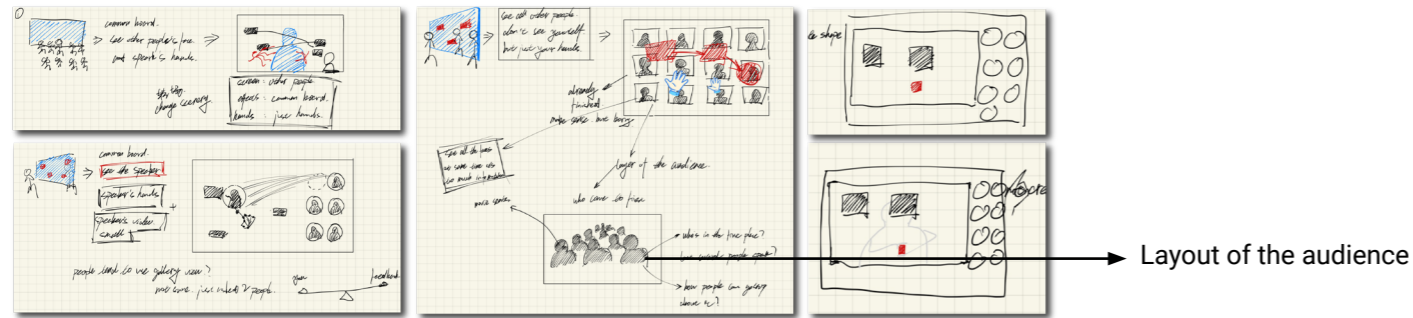
Hand Sketching

Aspects

Although it is hard to pick a single part to explain in the nonlinear process, I can still show some aspects I thought during the exploration, but again, all the ideas are connected and they should not be viewed individually.

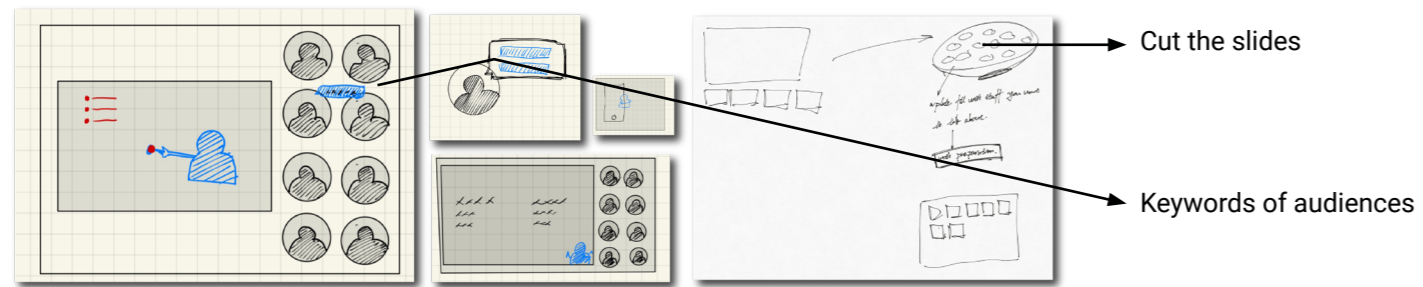
Platform

Most of my sketches are just the video of the speaker, but in the real situation, speaker always see the audience camera video. I need to put platform this factor into the situation.



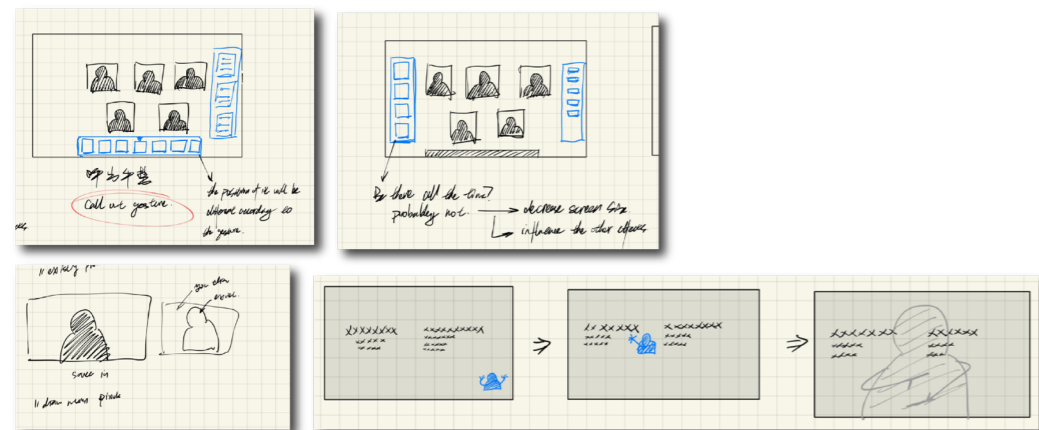
Sharing slides

How gesture interaction can help to explain ideas during sharing slides.

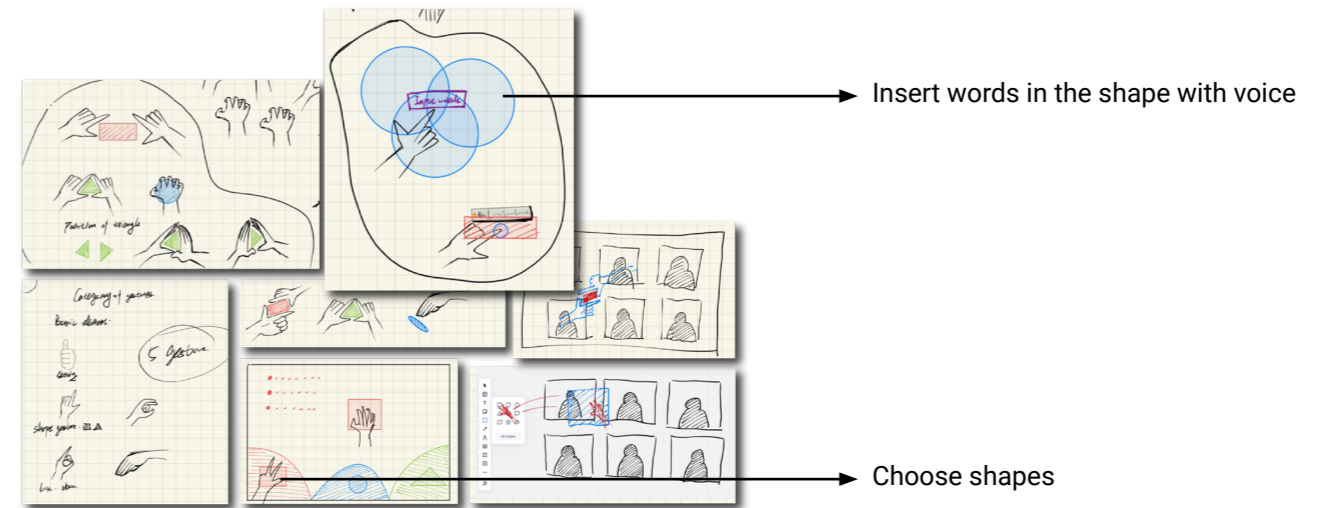


Change speaker and change topic

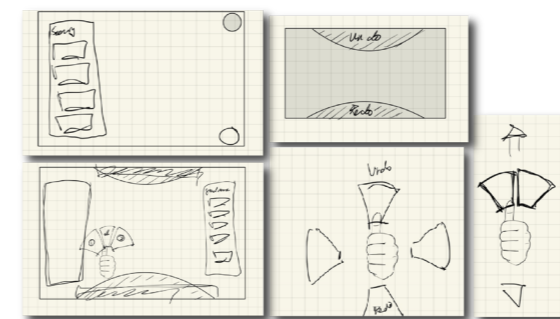
What will happen if the speaker is changed.



Different ways to create shapes



Undo redo

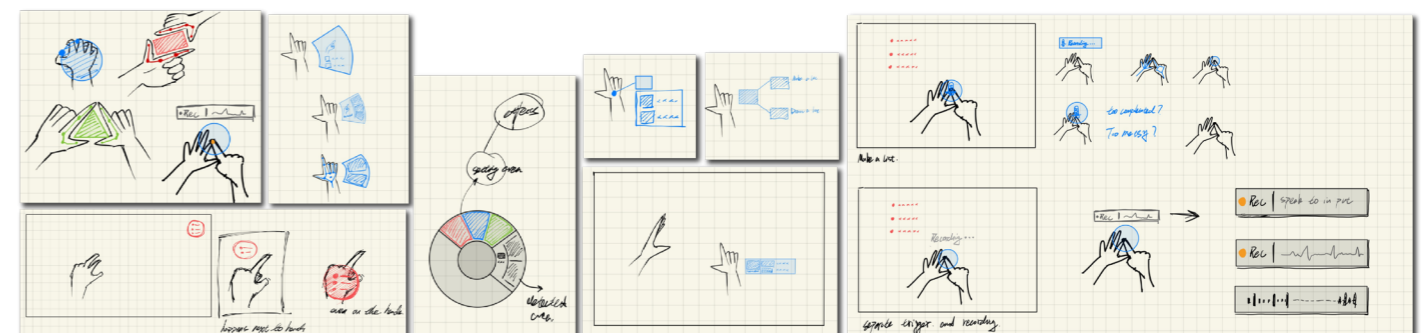


Mode - prevent mis-trigger



Indicate user

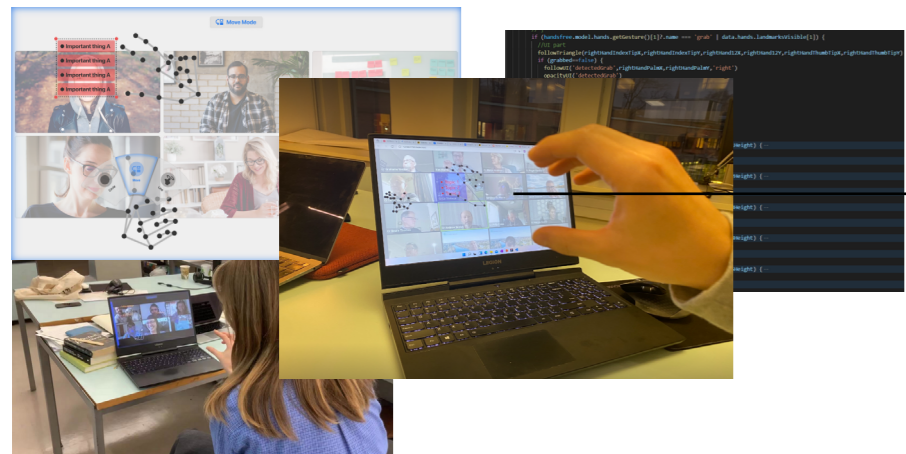
How to indicate to user what gesture has been detected, what's the status of the system.



Programming & testing

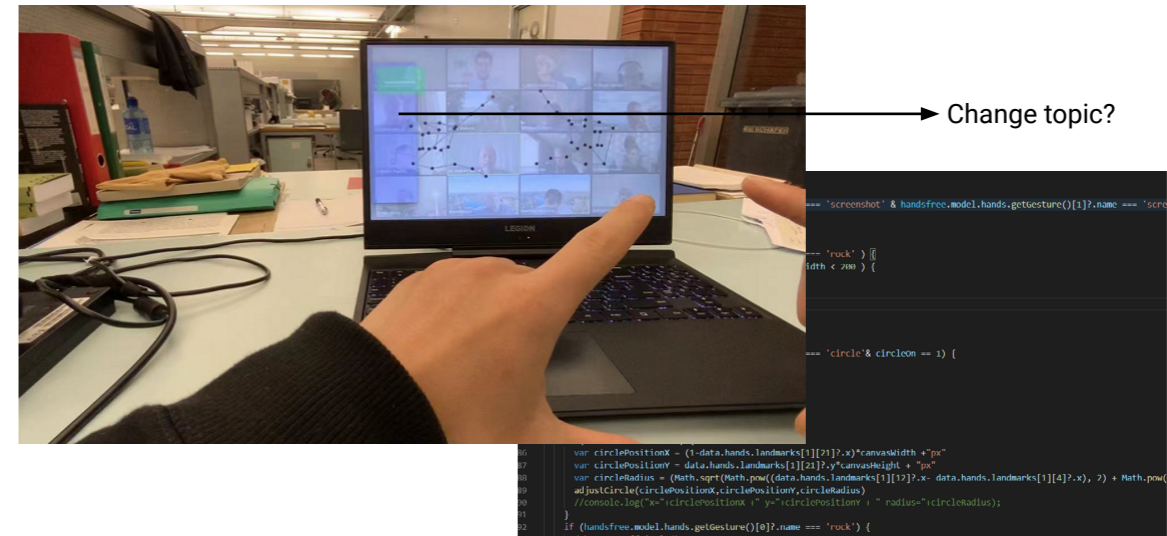
Some prototype and feedback

Grab and move the list



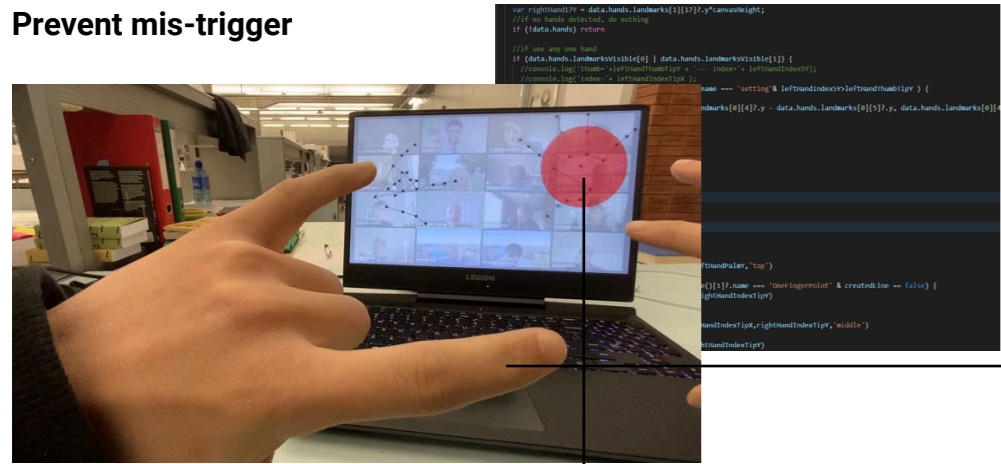
Grab a single word or whole list?

Move the topic



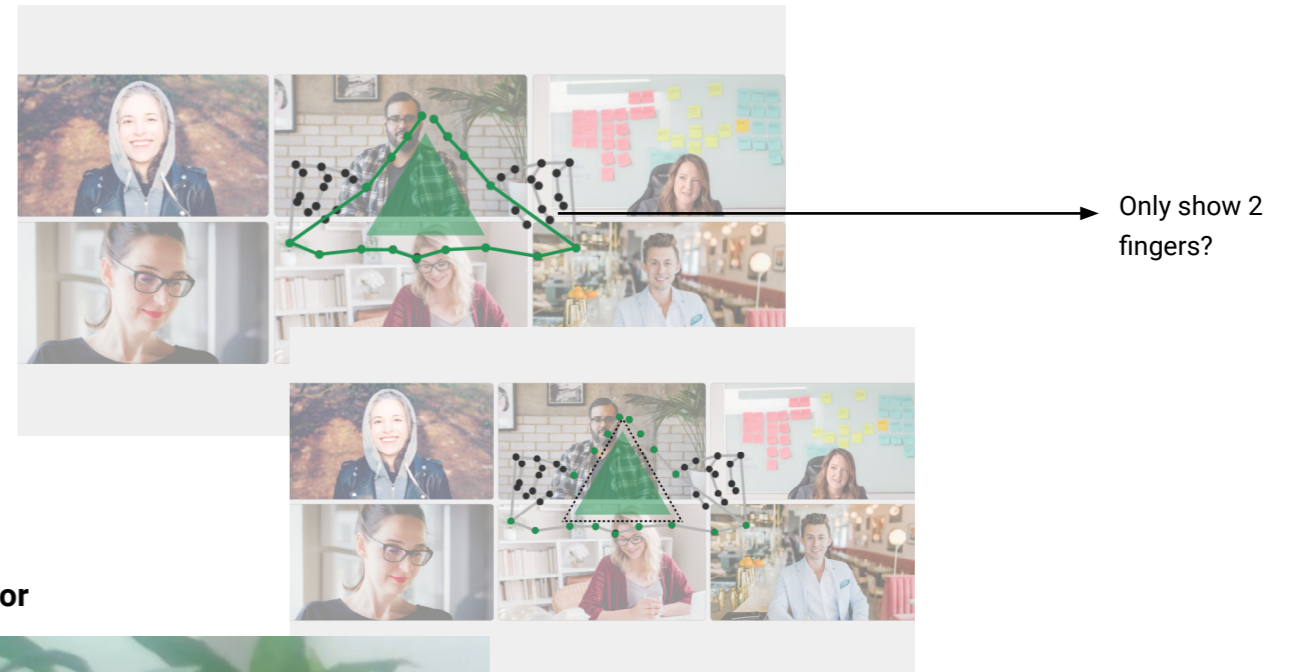
Change topic?

Prevent mis-trigger



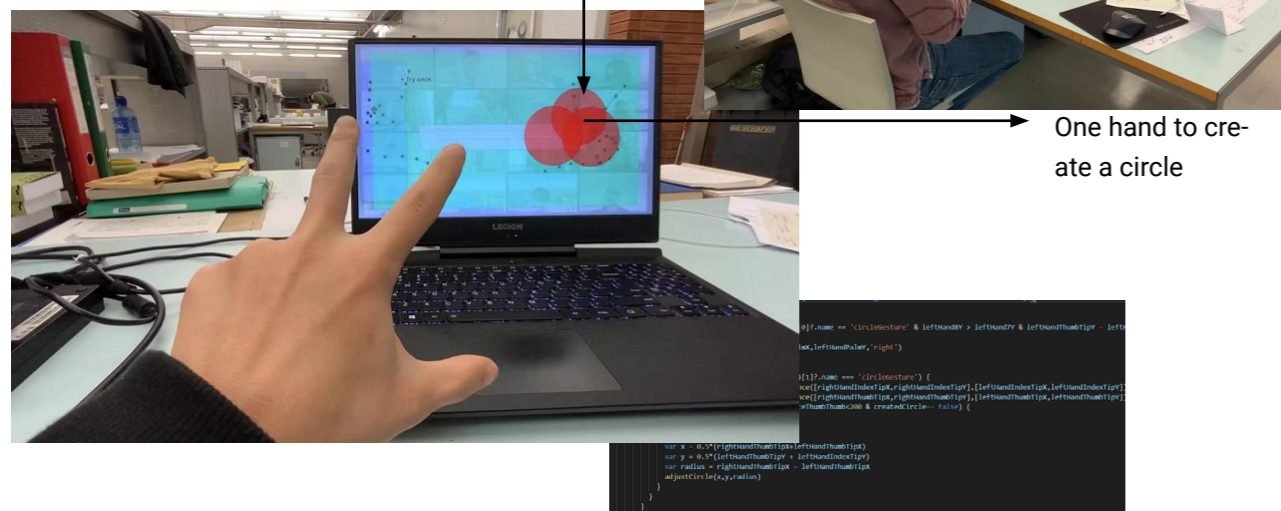
Very unique gesture to prevent mis-trigger?

Don't show whole hand?



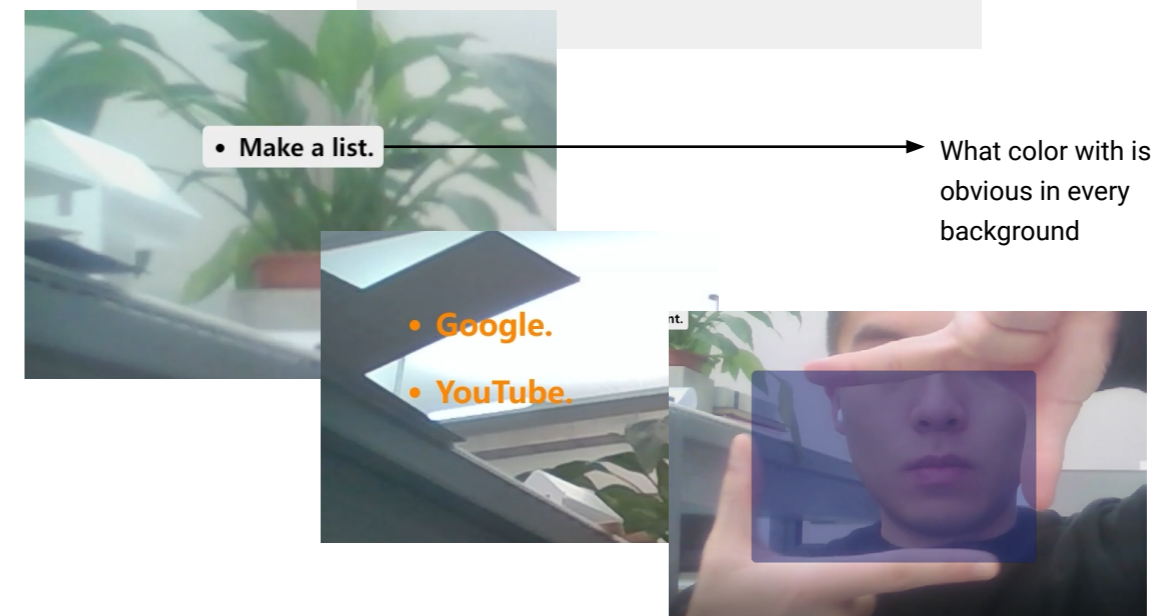
Only show 2 fingers?

One hand to create circle?



One hand to create a circle

Color



What color with is obvious in every background

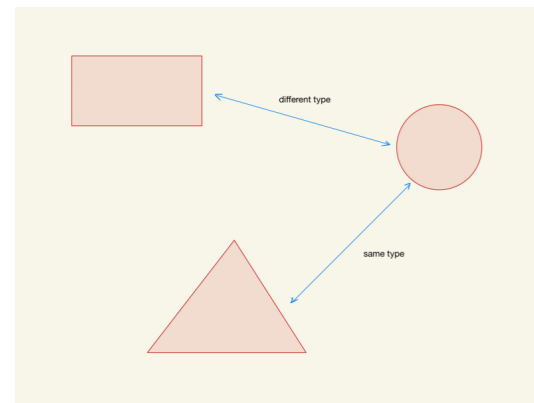
Extra research

As the exploration continued, some core functions such as making list, creating shapes have been determined, because of that, I need to conduct extra research to see how other people will use gesture to represent those functions.

How to gather natural gestures

Checking how other people represent shapes, making list, draw lines with gesture, the simplest way is asking them "how will you represent triangle with gesture?" However, this way is too deliberate, the gesture people give may be quite different with the one they really use in the communication. Because of that, I designed this way to gather gestures.

Gathering natural gestures of representing shapes



How

Ask candidate to explain this diagram without using the words "circle, rectangle, triangle, line" in 1 min.

Why

It is extremely hard to describe the diagram without using those words, especially in this short time, because of those limitations, candidate will use gesture more to explain it.

Gathering natural gestures of making list



How

Ask candidate to remember the list for 45s, then repeat it within 1min.

Why

It is hard to remember the whole list within such short time, but gesture can lighten speaker's working memory (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Wagner, Nusbaum, & Goldin-Meadow, 2004)., according to that, people will use more gestures during recalling this list.

Gestures such as representing shapes, lines, making lists can be gathered through the workshop above, however moving some virtual object is then not part of our natural gestures, I can only gather that gesture by asking people.



Gestures gathered



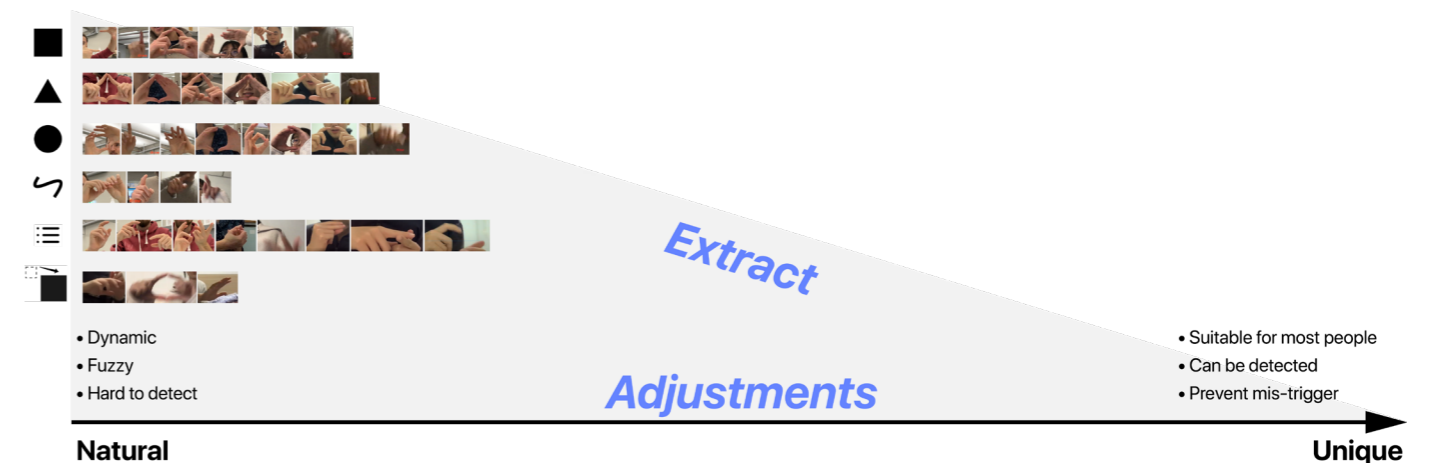
Insights - pattern

From this research we can see people's gestures are dynamic and fuzzy, but we can still find some pattern behind them. For example

- Most people use quite similar gesture to present triangle, circle, line.
- People mainly use index and thumb to represent the shape.

Insights - mis-trigger

Obviously using natural gestures would be the best solution, but natural gestures are also hard to detect, they are dynamic, fuzzy, personal. Instead of using the natural gestures directly, we should extract their features and make adjustments, so it can be recognized by the machine and also match with most people's habits.



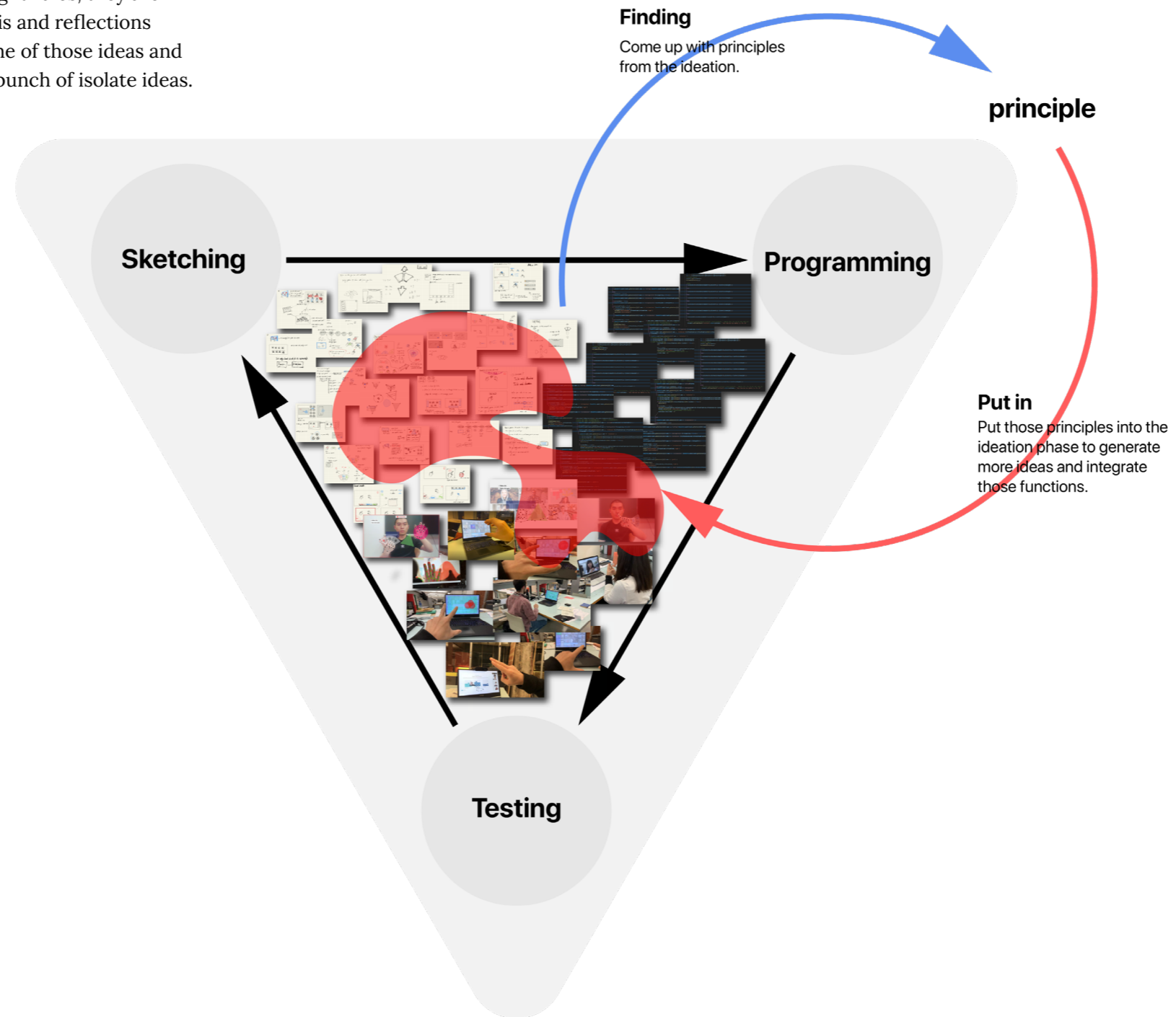
Principle

Hard to evaluate

Although I made many prototypes, but those prototypes are still far away from being used for usability test, the practical application of gesture interaction is not mature either, so there is no gesture interaction criteria that can be used to evaluate ideas, furthermore, as an exploratory project, there is no problem that I can check if it's solved. All those aspects make it hard to make decisions between the sketches.

My principles

As I mentioned before, during the process of the ideation, I also came up with some principles, just as the ideas, those principles are not the only right rules, they are based on my testing, the feedback I got from others, my analysis and reflections during the whole process. With those principles, I can pick some of those ideas and integrate them into one refined sketch, rather than showing a bunch of isolate ideas.



It's hard to explain those principles without video example, so they will be introduced after the refined sketch.

Refined Sketch

Based on the principles I came up with in the last stage, I made a more refined sketch. That's different with a final solution, those UI elements are low fidelity wire-frames, each function in this refined sketch can also lead to more possibilities.

Refined sketch

- Without slides
- With slides

Overview

- Core concept
- Structure of the functions
- 4 components + 1

Setting gesture

Modeless function

- Make a list
- Create shapes
- Manipulate blackboards
- Summary - prevent mis-trigger in modeless functions

Function with mode

- Draw lines
- Grab & move elements
- Summary -mode

Watch video then read text

Because gesture is dynamic, it's hard to explain by the static image or words, please click the hyperlink under the image or scan the QR code to watch the video. Watching the video first then viewing the text description, it will make more sense.

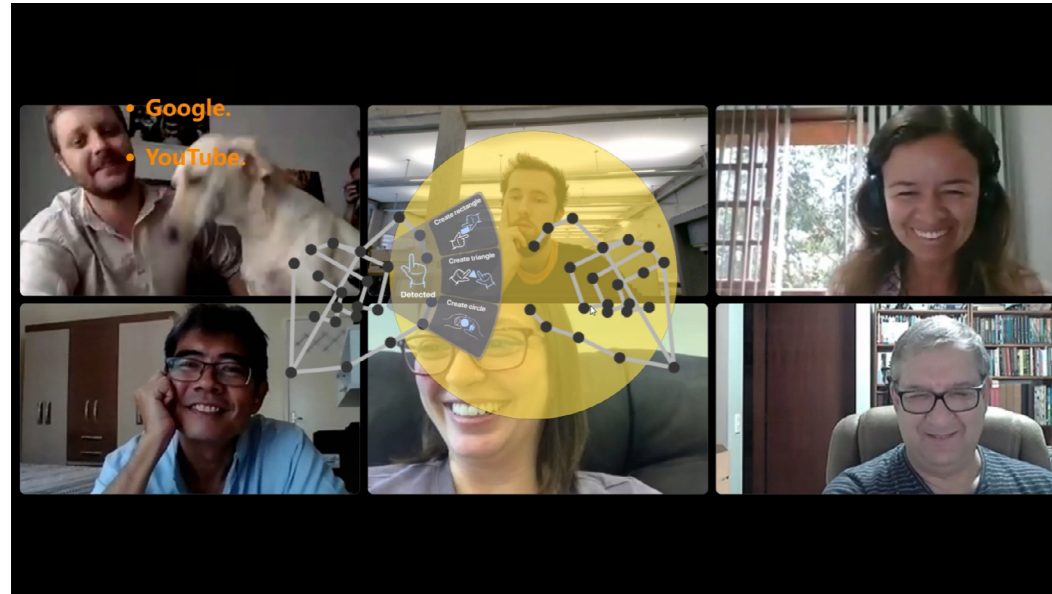
Experiencing it while reading? Here is my prototype

<https://gesturemeeting.netlify.app>

Refined sketch

- Without slides

Speaker's view



<https://ahocloud.box.com/s/7xoz0zb5nfvpi2ukalz6bjl4dqd19mo2>



The project involves the dynamic gestures and actions, please scan the QR code or click the hyperlink under the image to see the real effects

Audience's view



<https://ahocloud.box.com/s/vltjat0sn6ghb1f10sduqb3o3wiqyy4b>



Please scan the QR code or click the hyperlink under the image to see the real effects

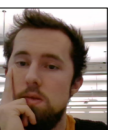


Wang

Role: **Speaker**

To explain the concept better, let's assume this Wang as speaker, the video above is from his view.

This is a standard desktop remote meeting software layout, you can only see other audiences' camera video in the meeting, but in this concept, you can't see yourself in the meeting except your virtual hands. By using the hands (gestures), the speaker can create lists, shapes to help him explain the ideas to the audience. The element he created will be placed into a transparent "blackboard" and the speaker can operate it.



Callum

Role: **Audience**

Assuming Callum as the audience, the audience will only see the speaker's video, which is similar to the speaker mode in Zoom. Furthermore, compare with the speaker's view, the interface element which the speaker sees is hidden from the audience, the audience can just see what the speaker created, he will just see speaker create the shapes, lists from hands.

Refined sketch

- Without slides

Audience's view



<https://ahocloud.box.com/s/7uvour0f41eui76g593dkqr6gt4bu90g>

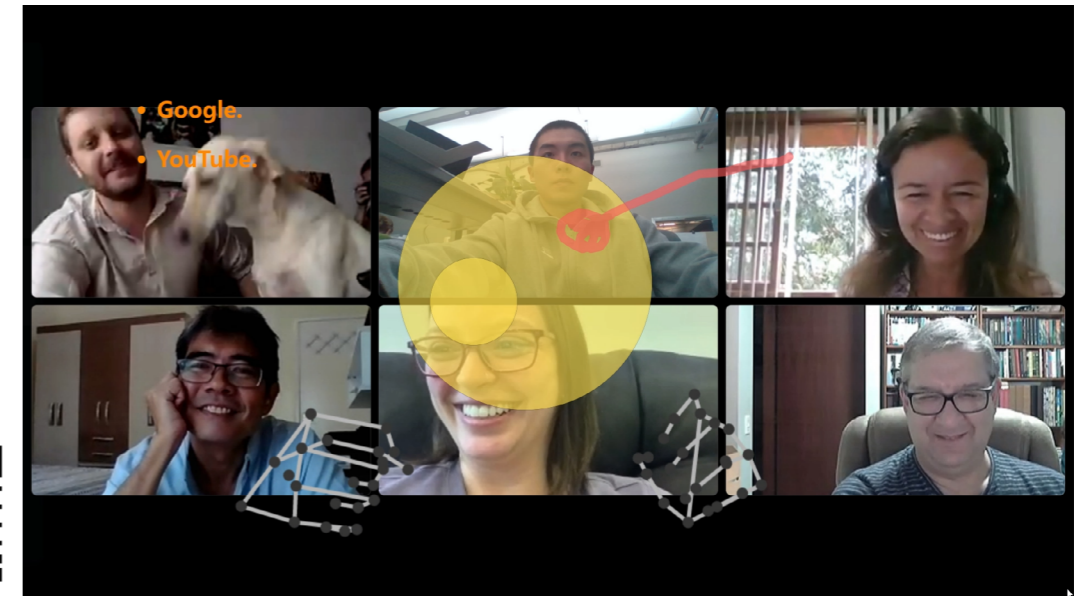


Please scan the QR code or click the hyperlink under the image to see the real effects.



Please scan the QR code or click the hyperlink under the image to see the real effects.

Speaker's view



<https://ahocloud.box.com/s/0t41g75i7exactuz3ef8qxtwiagfzyjd>

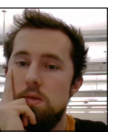
Change Speaker



Wang

Role: **Audience**

The video above is from Wang's view (current audience)
If Callum starts to talk, the speaker will change to Callum and Wang become the audience, as an audience, Wang's view is speaker's camera video.



Callum

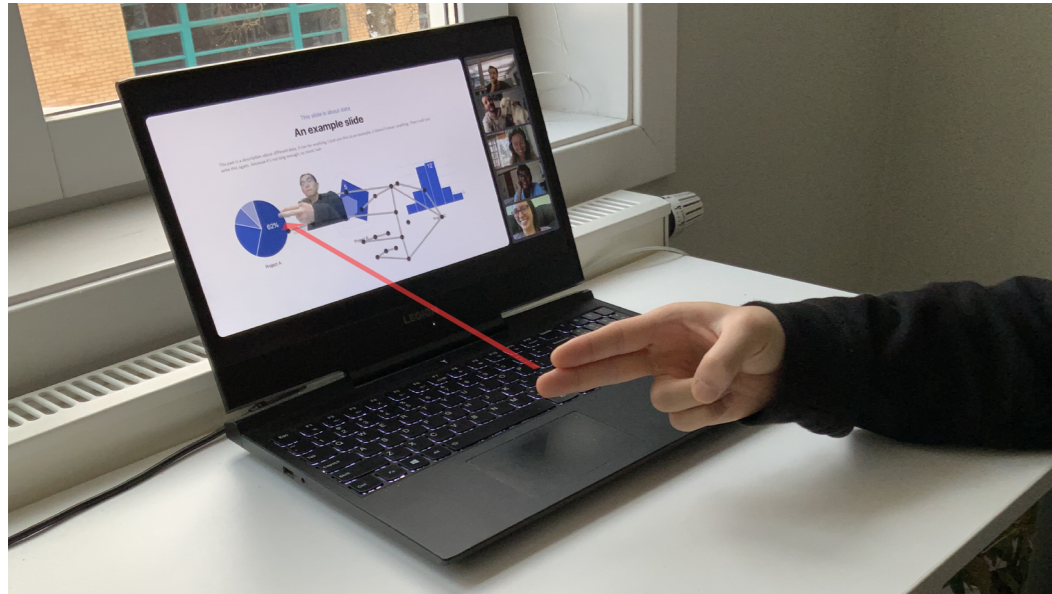
Role: **Speaker**

Now Callum wants to say something, he will become speaker and have access to the created elements as well as all common transparent boards, he can also create the new elements. Those transparent boards are for common, they don't belong to any single candidate, anyone who is talking can create things in the common transparent board

Refined sketch

• Sharing slides

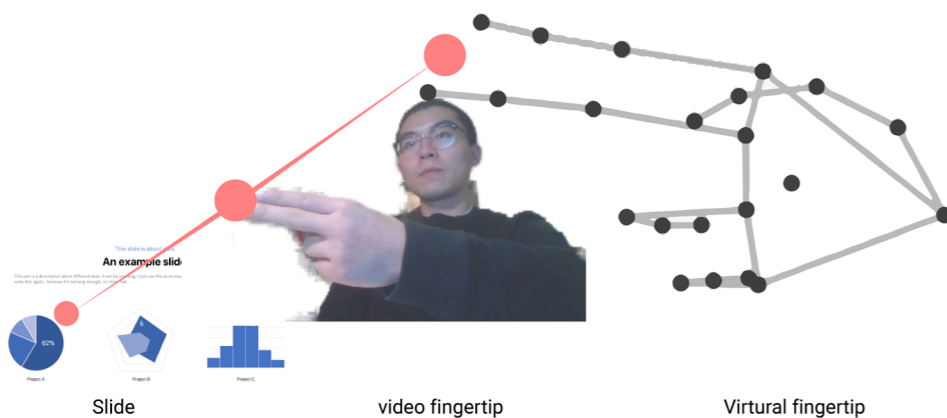
Point the place you really point



<https://ahocloud.box.com/s/6ny0vobm61huzp3kjaw0r6ics8ip73xb>

In the situation of sharing slides (or screen) in remote meeting, the basic function will be the same (making lists, creating shapes, grab, etc.) But there will be another additional function in this situation—pointing.

Instead pointing with a cursor, speaker can just point with their hands, and when the speaker point at something, his camera video will also move to the place he points. His video will also be slightly transparent to avoid hiding the content behind it.



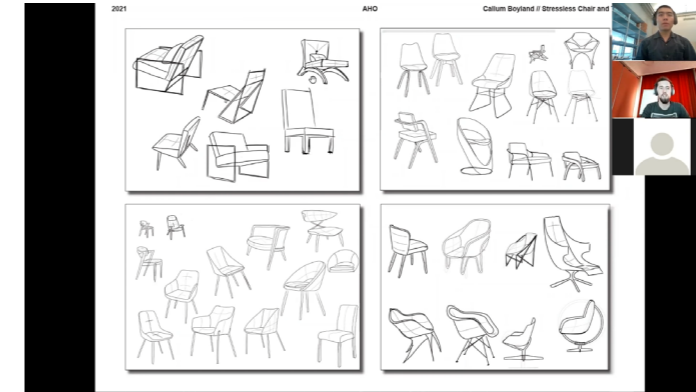
Furthermore, speaker's pointing gesture fingertip will be in the same place of his virtual hands pointing gesture fingertip. By this way, we can simulate the same effect as in physical presentation speaker pointing at the place he's talking about.



Please scan the QR code or click the hyperlink under the image to see the real effects

Why?

Some insights



Point with cursor



Point with hands

From the research before we know that “pointing” is actually deictic gesture and it's for directing audiences' attention, this comparison of pointing with cursor and pointing with real hands shows that pointing with real hands draws much more attention.

I assume the reasons of this are

- Speaker's hand is bigger than cursor.
- As human, other audiences are more sensitive to the hands

In real life talks, audience attention bounces between the speaker and the slides. In a way, there are two characters on stage. You can play games with that.

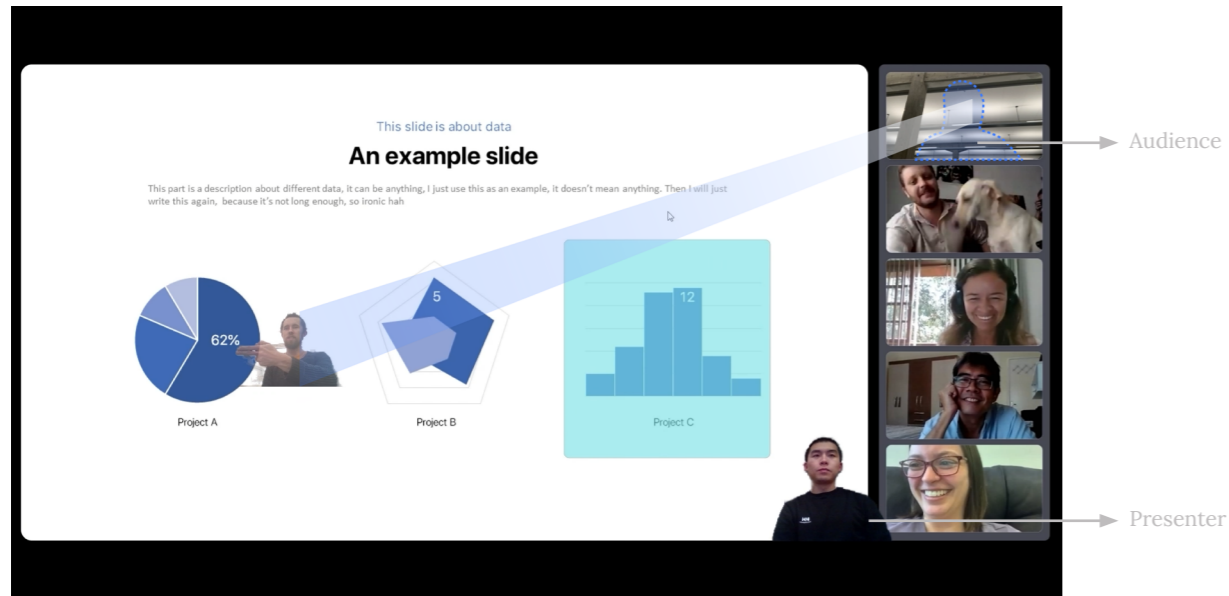
https://interconnected.org/home/2020/05/15/video_talks

I quite agree with Matt Webb's opinion that audiences should have opportunity to switch their attention between slides and speaker in remote meeting. I think that's a solution to prevent people from getting bored or being distracted.

Refined sketch

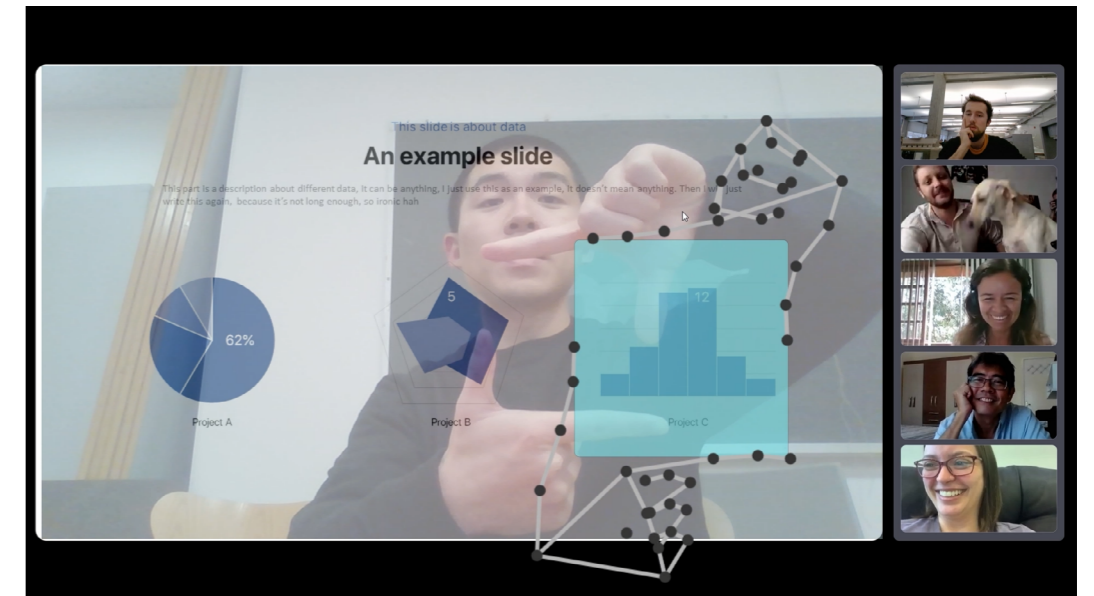
• Sharing slides

Audience's pointing



When presenting, there is a main presenter (the one who shares the screen), but the audience can still use the pointing function, when the audience points, his video will also appear on the slides.

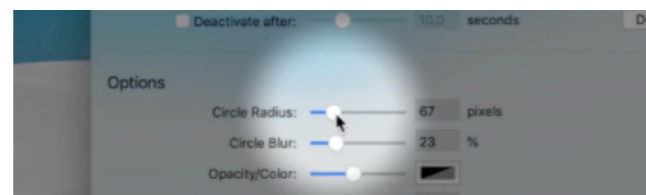
Other functions



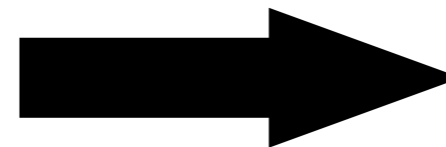
Except the pointing function, the other functions are the same in the screen sharing situation, the speaker not only can create shapes to explain the spatial relationship, but also can use those functions in their special way, for example, highlighting part of the slide.

Reflections

As we all know that our attention is very fragile, people may doubt if this video moving on the screen can help us focus or just take away more attention from the place we point. Because of this very limited time and my poor development ability, I cannot conduct experiment to test it (it's a working prototype, but it's far away from conducting usability test, for example, in my prototype, the video is shaking because the program keeps detecting the gesture from camera, it can be easily eliminated by some professional developers)



In my assumptions, comparing with the typical way of pointing, which forces audience's attention to the place where speaker want to point (image above), this solution is more similar to our physical presentation, it does add another character on the stage, but it can help avoid audiences getting bored, hence paying more attention to the slides in total.



Potential testing

If there is a more advanced prototype, maybe we can have this single factor experiment:

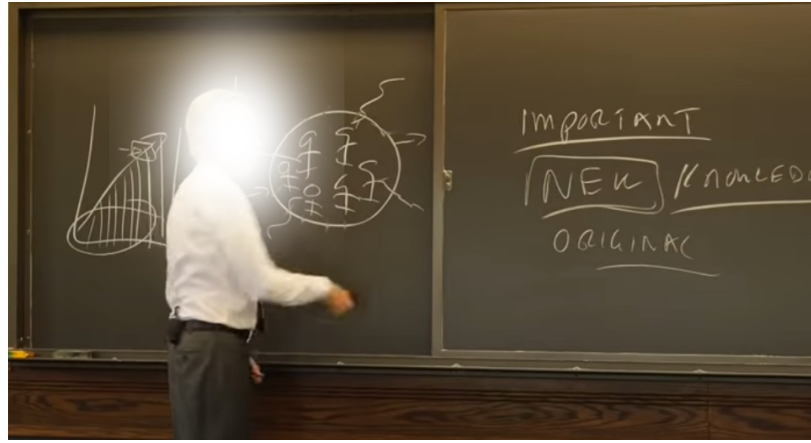
1. Prepare the presentation which holds some data (random number) in each slide.
2. Prepare random candidates and divide them into 2 groups as audience (Group A, Group B).
3. Use the same speaker presents the presentation for 2 groups, Group A, speaker point the number with cursor during the presentation, in Group B, speaker point with this point function. The presentation should last about average presentation time, for example 30mins.
4. Ask candidates in Group A and Group B to repeat the data in presentation, check which group can repeat more and correct.

Refined sketch

Overview

- Core concept
- Structure of the functions
- 4 components + 1

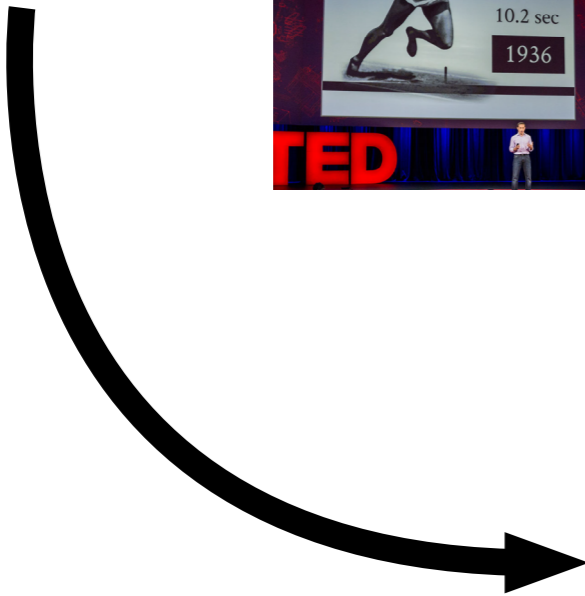
Inspiration



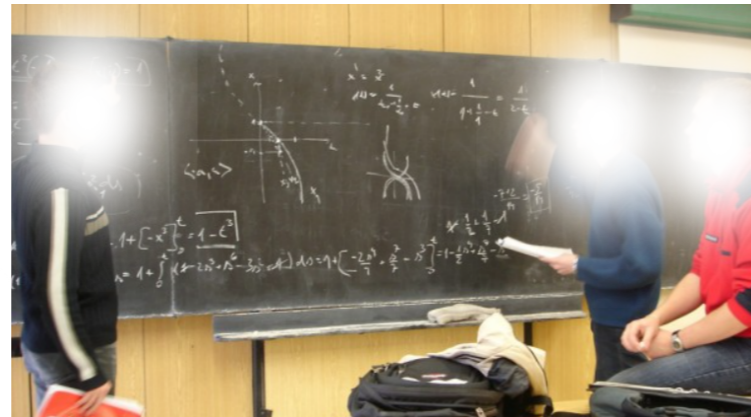
The inspiration of my project comes from the blackboard in classroom, a classic way of explaining ideas, the teacher (speaker) will use simple words, shapes to support them expressing ideas.



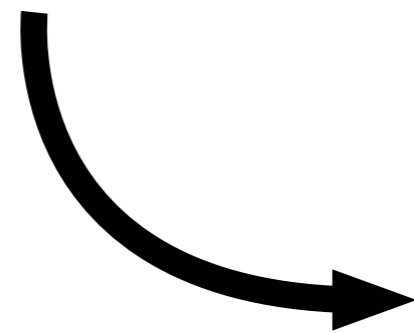
Compared with using slides, writing down words or drawing shapes happens while talking, it's real-time and normally accompanied with the speech, just like our gestures.



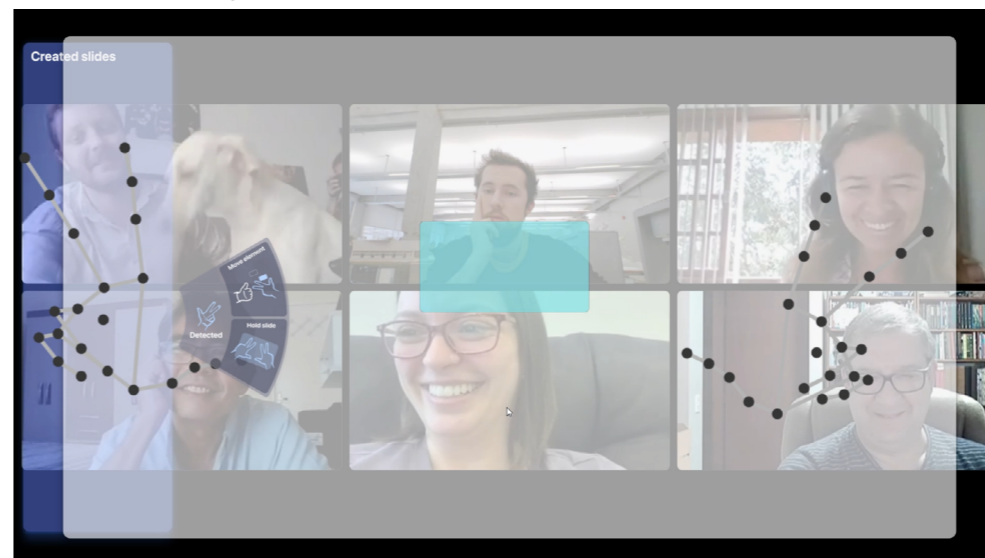
Difference



Unlike the normal classroom, in this project, the concept is that every candidate of the remote meeting faces the same transparent "blackboard" and there is no certain host, everyone can create elements in the transparent "blackboard". It's a common transparent board



My concept



The common board in the concept will be transparent normally, so the speaker can still see all other audiences' faces, when the speaker uses gesture to hold the common transparent board, it's then solid, because of this, it's easier for speaker to change the board or grab it back.

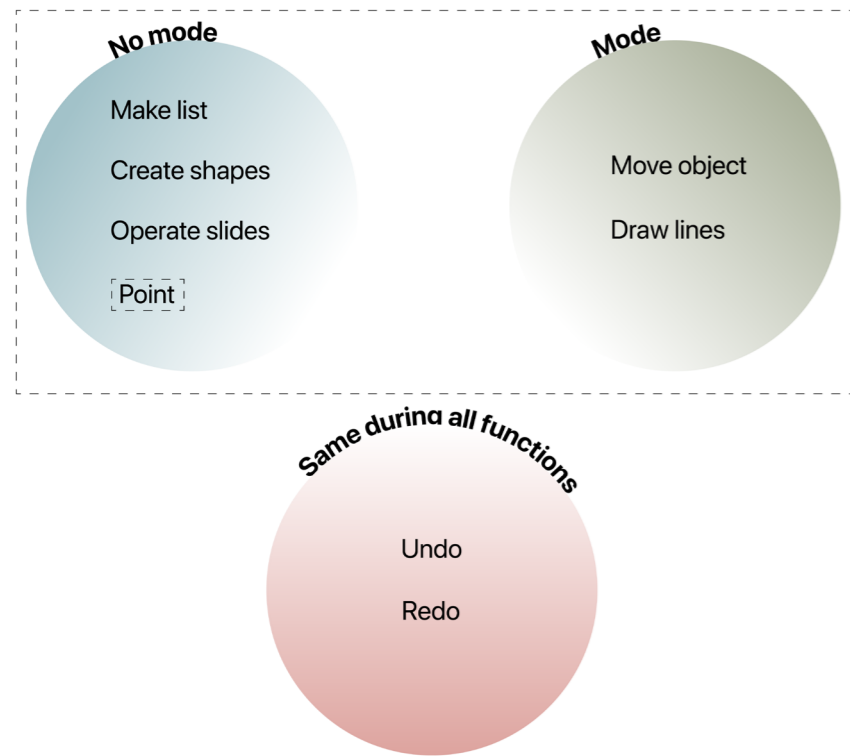


Please scan the QR code or click the hyperlink under the image to see the real effects

<https://ahocloud.box.com/s/urs02g27jei0drhx3z3c2zv6g5pmyja3>

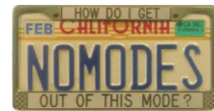
Core concept

Structure of the functions



Mode & No mode

All functions showed in the concept can be categorized into those three groups. As mentioned before, putting those functions into some mode can help prevent mis-triggering problems, but entering mode itself is an extra step, it's more like a compromise between preventing mis-trigger and using more natural gestures.

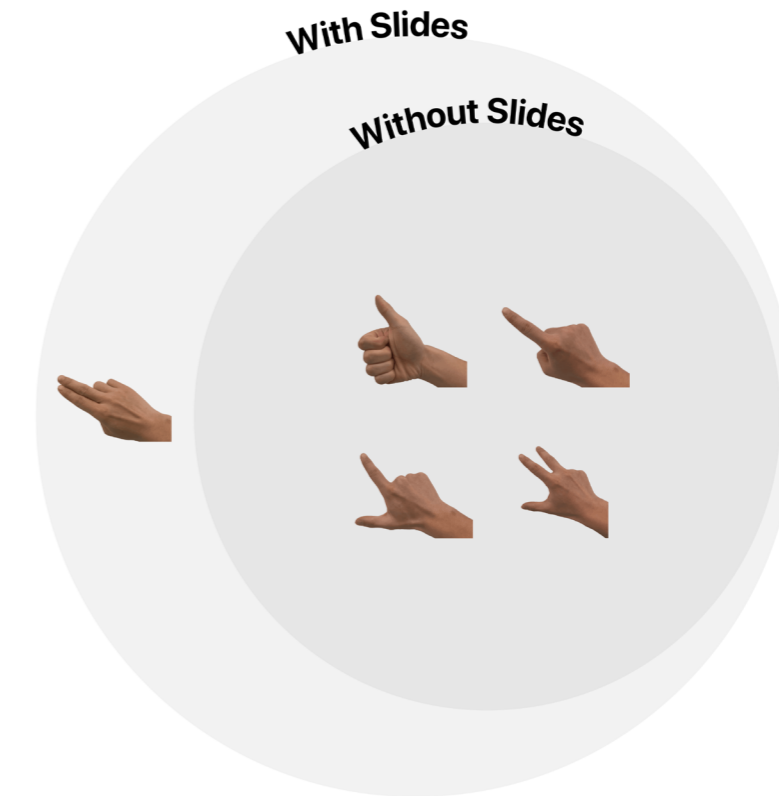


This mode concept also reminds me how Larry Tesler wants to avoid using modes, which I really agree, again, mode here is a compromise between preventing mis-trigger and using more natural gestures.

A mode is a context for getting work (or play) done. In the early days of computers, before graphical user interfaces, applications were broken into "operational modes"

<https://www.pavley.com/2020/02/20/no-modes/>

4 components + 1

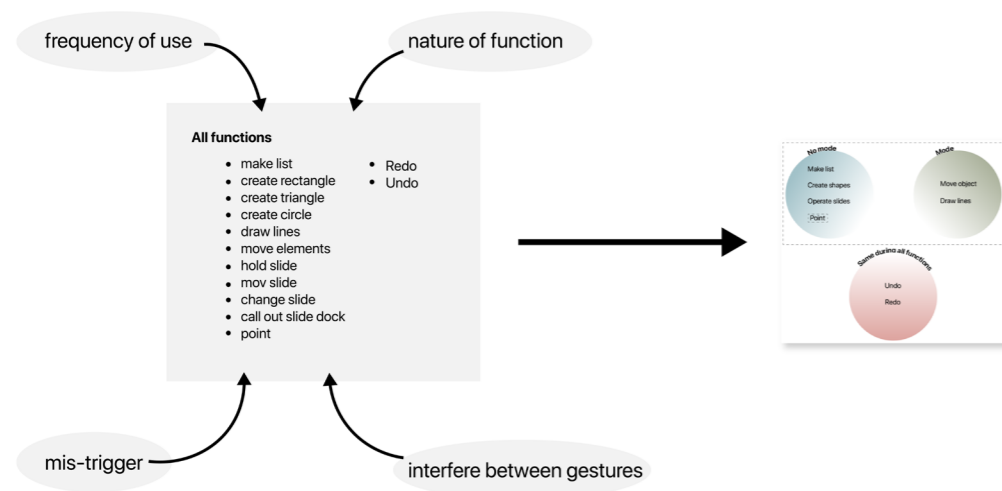


Less gestures

There are many functions in this concept, since the target user is a regular user, there should be some way to make it easy for them to use. The first thing is to limit the number of the gestures which trigger the function.

4 components + 1

In total, there are only four gestures in both with and without slides situation, and there is one additional gesture only for pointing function when sharing slides. In stead of calling them gestures, components would be a more suitable word, because most functions require the combination of more than one gesture to be achieved.



Why

Grouping those functions in this way is a decision that considered many factors like frequency of use, nature of function, mis-trigger, interfere between gestures...

Refined sketch

Setting gesture

Setting gesture

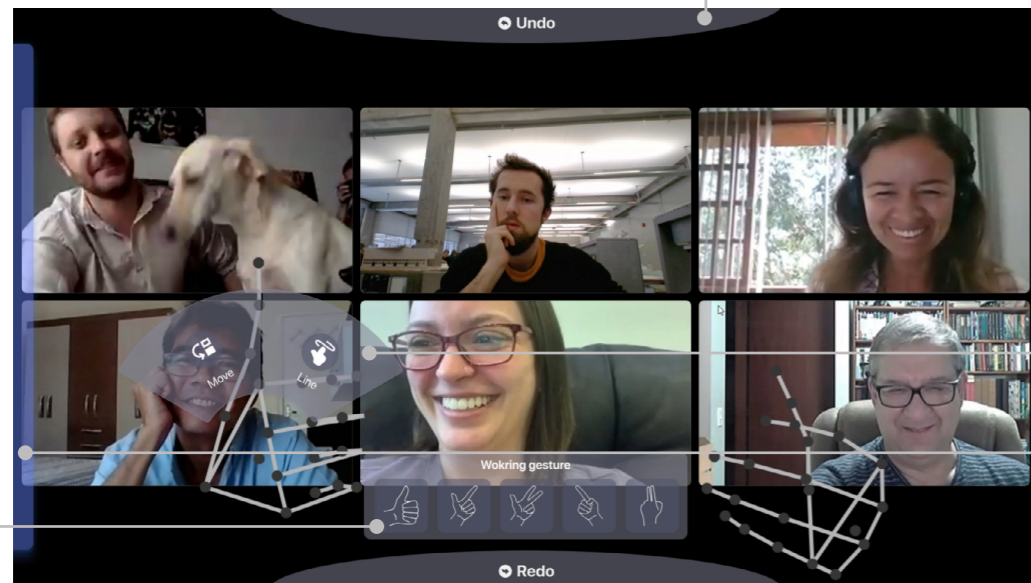
Before going through the details of each function, I need to start with this setting gesture.



Setting gesture

Setting gesture here is a functional gesture without any meaning, it serves to check all the components, choose modes, trigger some function zone in the screen or prevent mis-trigger problem. It is like a dashboard on your hand, and you can call it out by this setting gesture.

Call out the dashboard



Please scan the QR code or click the hyperlink under the image to see the real effects

<https://ahocloud.box.com/s/5q9dfzsq7lrlrb1hev460eqgxeq5jja0>

By representing the setting gesture, all functional areas and tips will pop up, those areas can remind user what they can do by gestures in this project.

What component can be detected



This area will indicate what gestures/components can be detected, technically users even don't need to remember all the 5 components, but just the setting gesture, and then check this area. By indicating user what can be detected here, then combining with indicating users what they can do with the detected gesture (it will be mentioned later), user can use all the functions with just remembering the setting gesture.

Undo & Redo

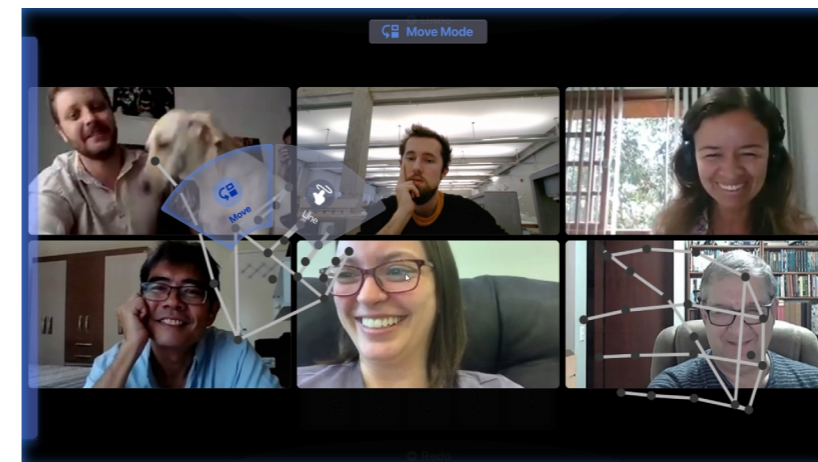


<https://ahocloud.box.com/s/k7l2w6d93fo2uhr5kvlgycgwvti5lveq>

Please scan the QR code or click the hyperlink under the image to see the real effects

Keeping this setting gesture and moving into undo or redo trigger zone, then the undo/redo function will be triggered. As mentioned before, it is same during any situation.

Choosing mode



<https://ahocloud.box.com/s/0te2i7joapnd9z4xuuvwvlw6jeobnabr>

Please scan the QR code or click the hyperlink under the image to see the real effects

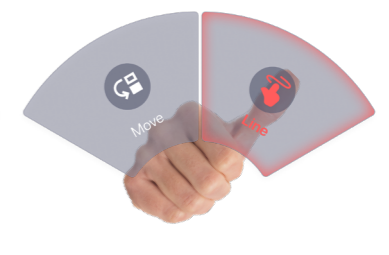
Together there will be two modes, move mode and line mode (it will be introduced in detail later), every time user represents the setting gesture, there will be a mode choosing menu follow his gesture hand, user can use his thumb as pointer to choose different mode.



Not choose

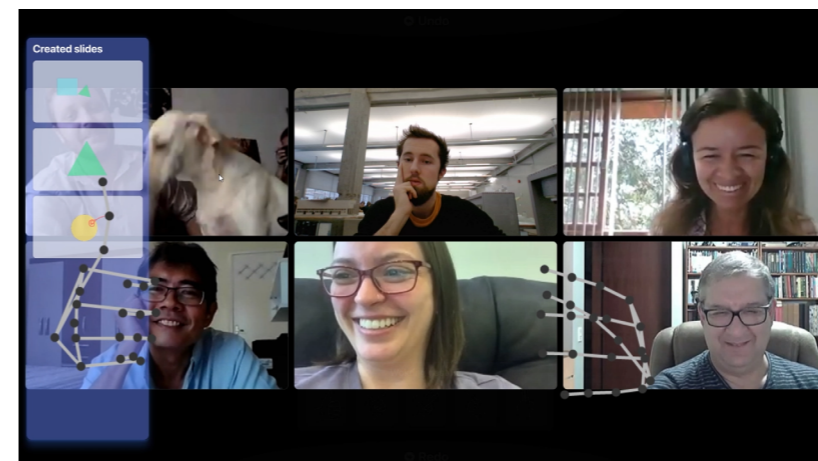


Choose move mode



Choose line mode

Call out transparent board dock



<https://ahocloud.box.com/s/oqcla6h7m6n6ex0wf18ojhs4hf6b7mqn>

Please scan the QR code or click the hyperlink under the image to see the real effects

Using setting gesture to touch common transparent board dock area can call out all the boards people created, using the same gesture to go out of the area, the board dock will slide back.

Refined sketch

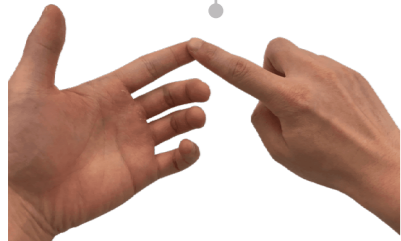
Modeless function

- Make a list
- Create shapes
- Manipulate blackboards
- Summary - prevent mis-trigger in modeless functions

Modeless Function-**Make a list**

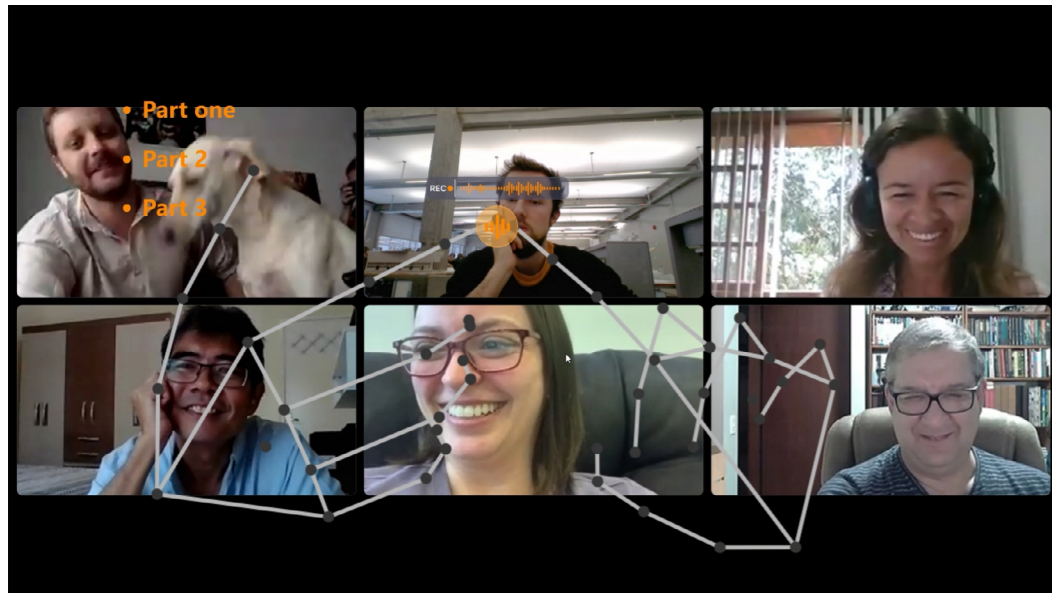
After introducing the setting gesture, let us go through all functions by 2 groups, first group is those functions without mode. Let us start with making list.

Used component



Make a list

Using the gesture above then using index fingertip to touch another hands' fingertips to make a list.



<https://ahocloud.box.com/s/cig2x68py1r8vb557rlo6mjumztupuk7>

How it works

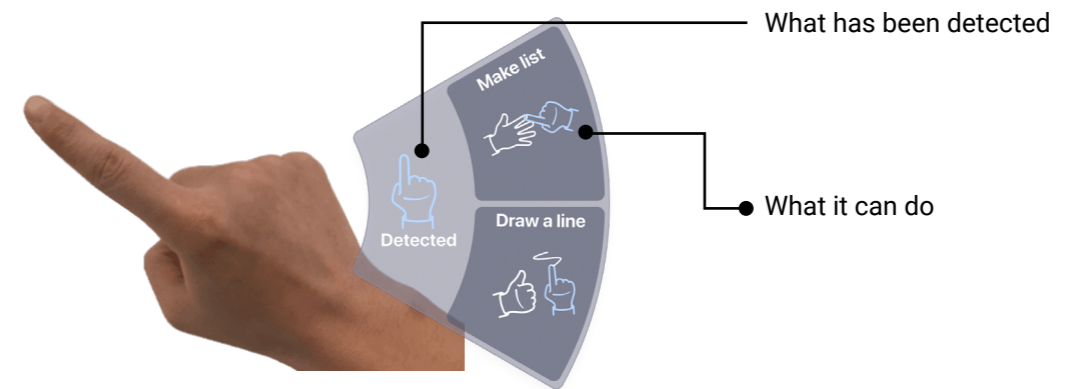
When the certain gesture has been detected and two fingertips is overlapped, then the machine starts to record people's talking and transfer the speech to text, after that, put the text on the screen.

Time order

In my prototype of this function, the user needs to perform the gesture to trigger the function first, then say the words. However, in real life people normally perform the gesture and speak at the same time, then that can be a problem. But this issue can be solved by recording all the way in the backend, when user performs the gesture, we show the words 1 second ago.

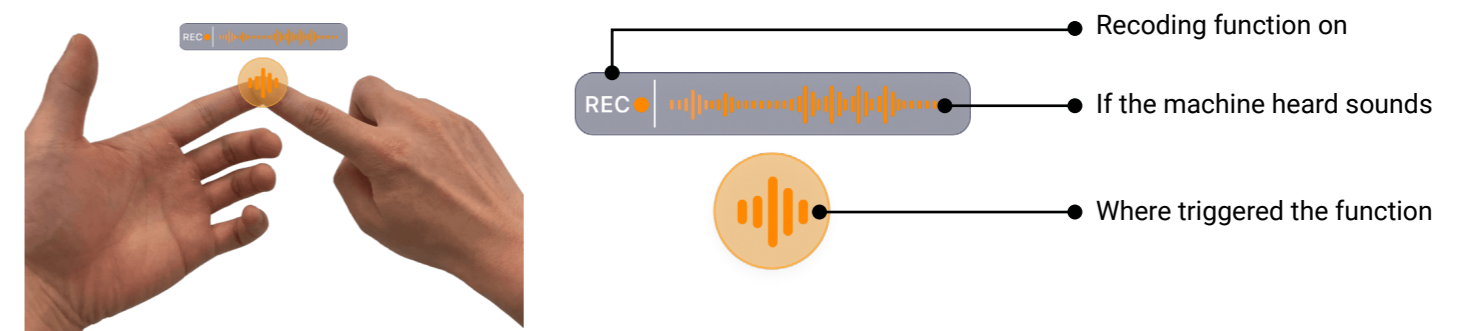
Indicate detection status

Every time the component has been detected, it will indicate user and tell the user what he can do with this component (gesture). This Indication UI shows up next to the hand which perform that gesture, because that's the place the speaker will pay attention to.



Indicate status of system

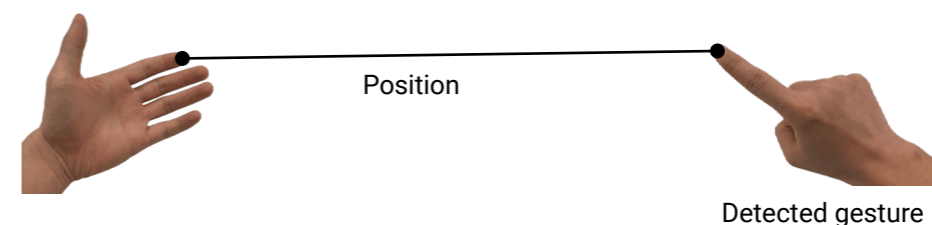
When the "making a list" function has been triggered, there will be another indicator to show user the status of the system, in this case, "recoding function on." And it will also show up at the place where the triggered the function.



Prevent mis-trigger

As a function that will be turned on all the time (user can trigger it without entering any mode), the critical thing is to prevent mis-trigger problem. In this case, it avoids being mis-triggered in this way

Not only the certain gesture needs to be detected, but also the position of certain fingertips of two hands should be in the same place.



Modeless function-Create shapes

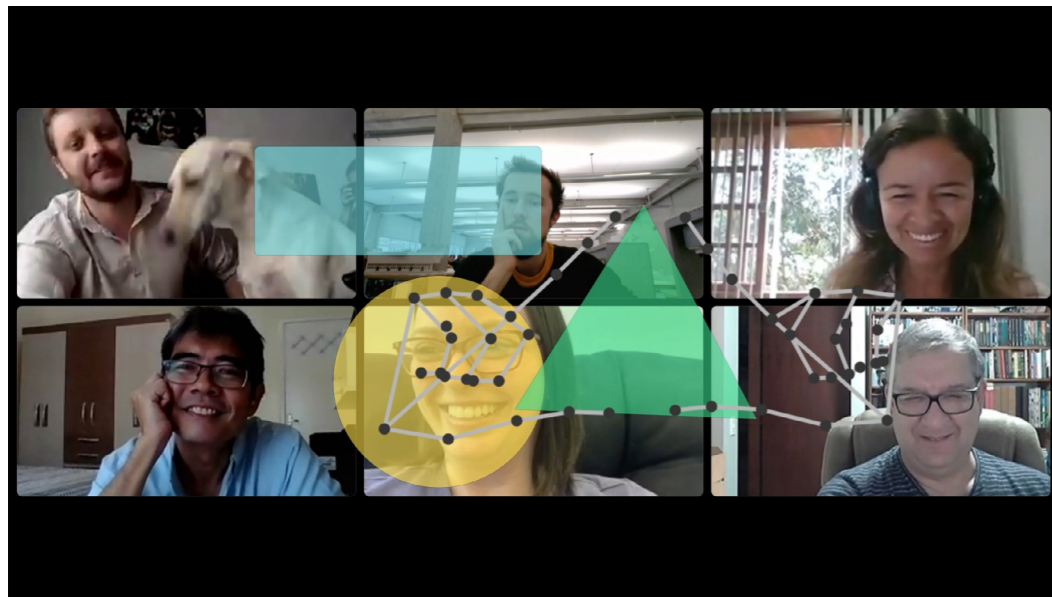
Another function that does not require entering mode is creating shapes.

Used component



Create shapes

By using the same component in different way, user can create different shapes: rectangle, triangle, circle.



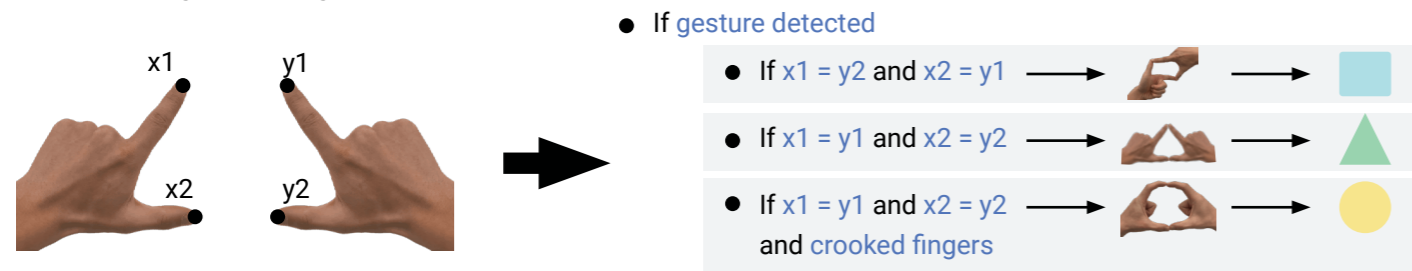
<https://ahocloud.box.com/s/rmir4yn5usns4sa7yfkt84n4ae0e7lad>



Please scan the QR code or click the hyperlink under the image to see the real effects

How it works

The principle of this function is the same as making list, but there is an additional conditional expression in this function, by calculating the positional relationship between two hands' thumb and index fingertips, it can detect if user wants to create circle, rectangle or triangle.



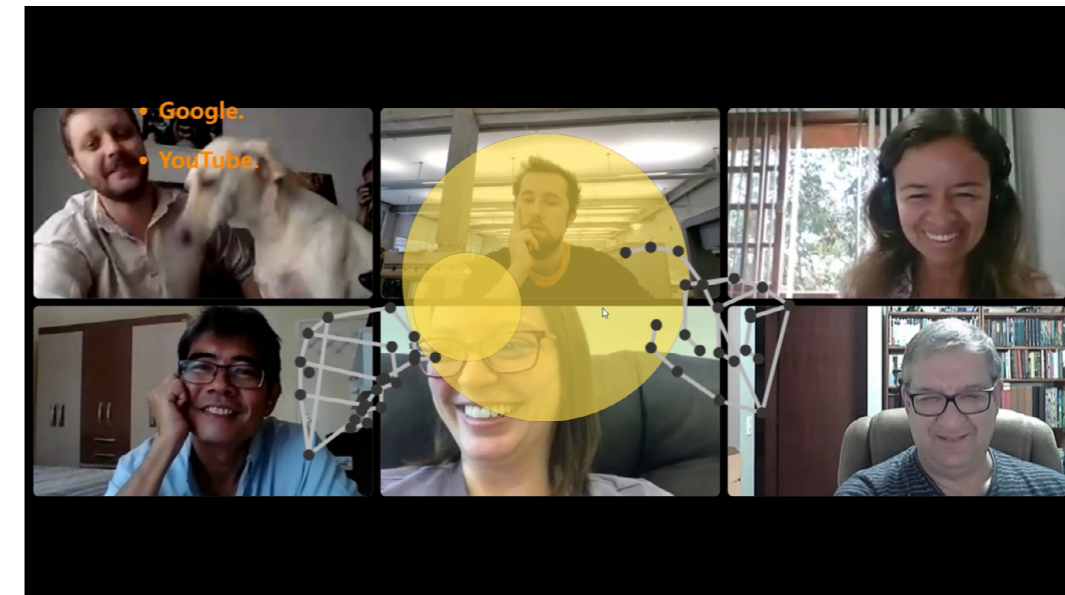
Same aspects

- Indicate detection status
- Indicate status of system
- Prevent mis-trigger

These three aspects are similar to those of making list

Why creating shapes

Creating shapes itself is not the purpose, those shapes serve to help us explaining ideas, especially spatial relationships.



<https://ahocloud.box.com/s/m8tsf2ya28cysrrcnesslmmrsg1rlmdd>

Except circle and overlap relationship, we can also use triangle to explain the hierarchy, those are just examples, there is a lot of possibility. When sharing screen, we can even use shapes to highlight some area just as what I showed in the beginning.

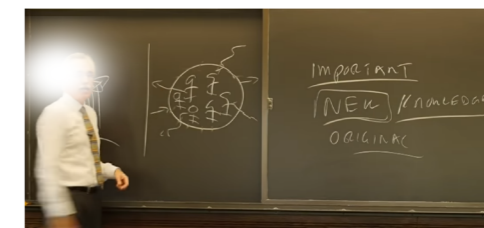


Please scan the QR code or click the hyperlink under the image to see the real effects

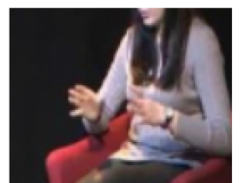
More reference of how people using shapes to explain ideas



My supervisor uses simple shape to explain ideas



Teachers draw shapes on the blackboard

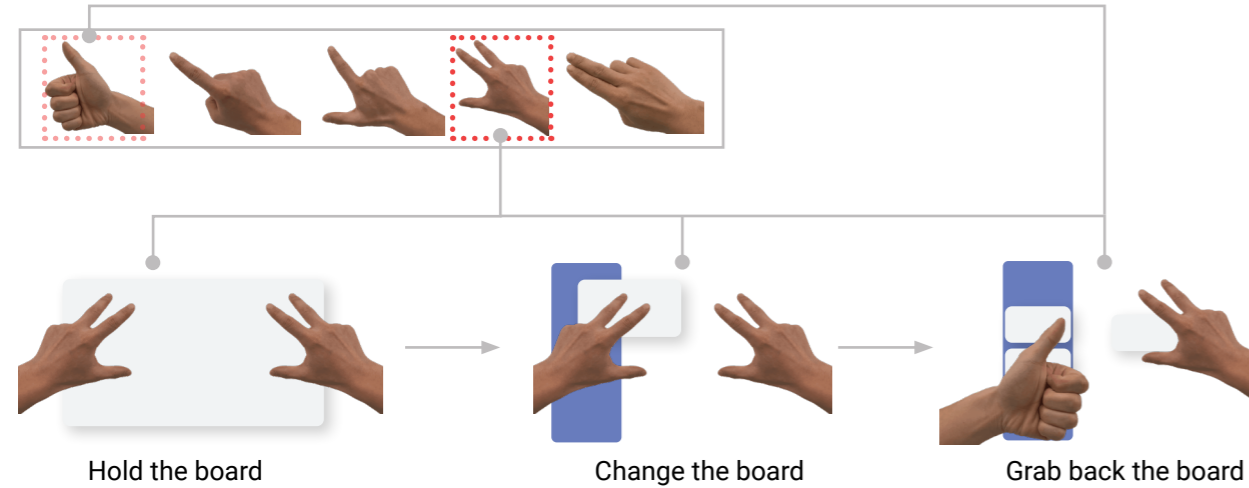


From the research before, our nature gestures serve to represent shapes and provide visual information.

Modeless function-Manipulate transparent board

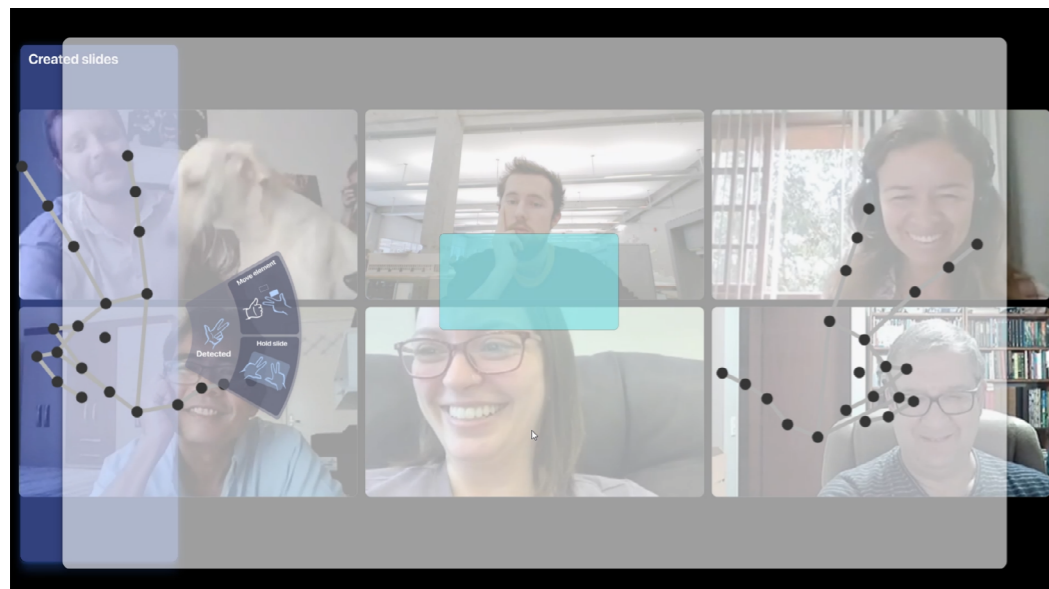
The last function without entering mode is manipulating the common transparent board

Used component



Hold & Change the common transparent board

With both hands perform the component, user can hold the current board, then if user uses one hand move to board dock (keep the gesture), it will move the current board to the dock and leave user a new empty board.



<https://ahocloud.box.com/s/urs02g27jei0drhx3z3c2zv6g5pmyja3>

Grab back

User can use the setting gesture touch the dock and call out the dock, then use the component (same as holding the blackboard) to grab the board back, with this function, user can go back to the topic being explained. Besides that, other candidates of the meeting can also grab back the old boards and discuss around it.



<https://ahocloud.box.com/s/fpryiyzxxdhcbnina9139wyqwh9hjr6>

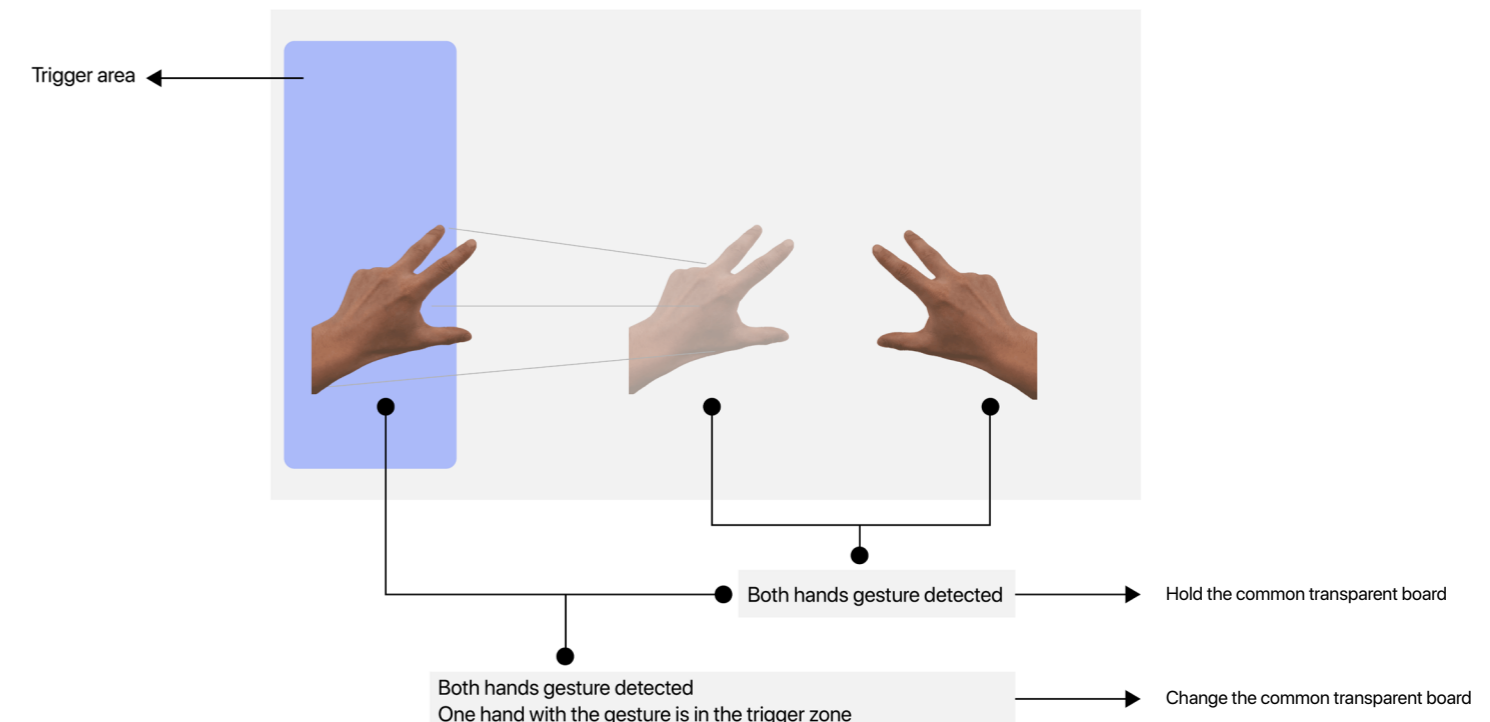
Although user needs to use setting gesture to achieve the function, but he doesn't enter any mode, the setting gesture only serves to call out the dock. More details about grab this action will be discussed later.



Please scan the QR code or click the hyperlink under the image to see the real effects

How it works & prevent mis-trigger

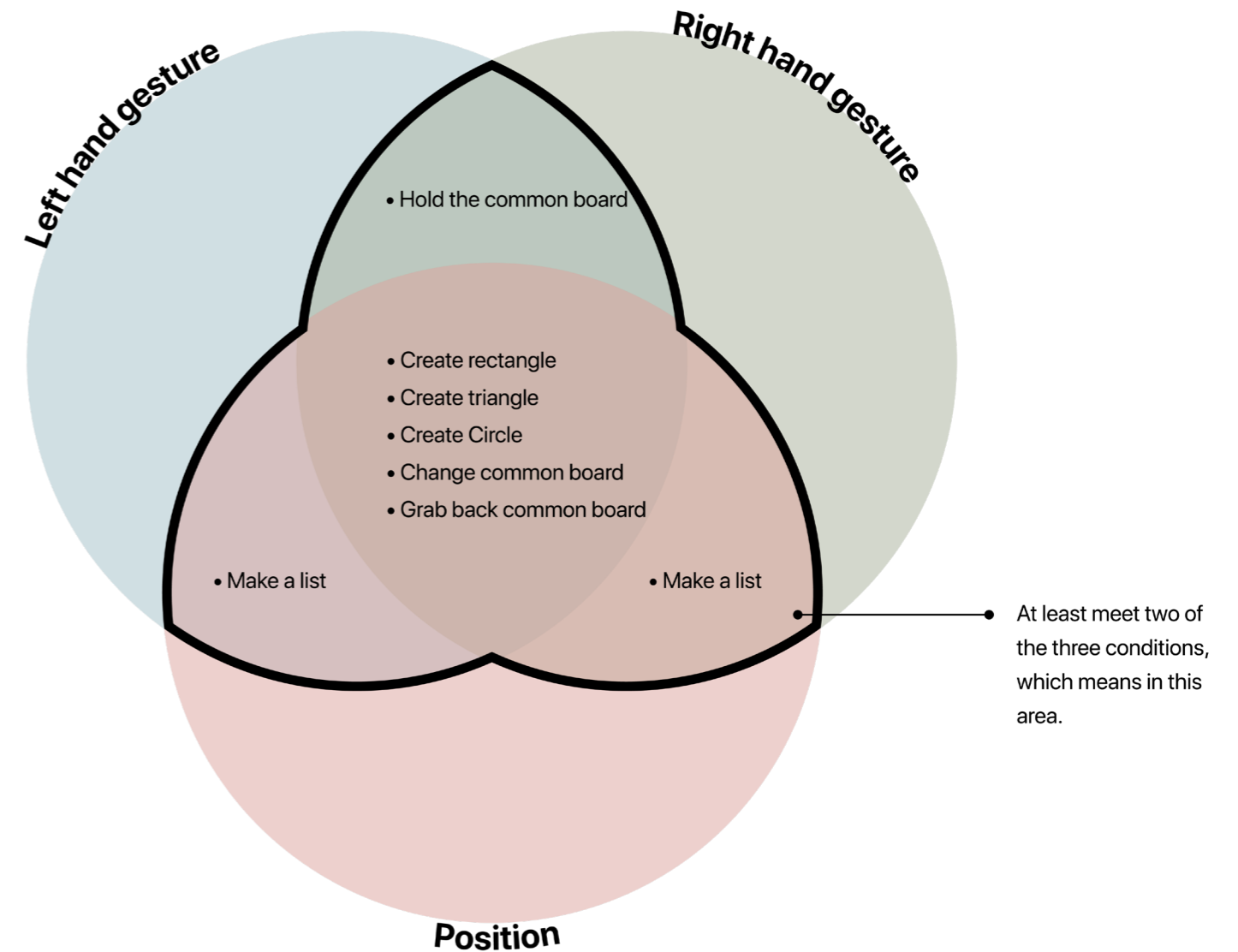
Since the trigger condition of this function is relatively simple, how the function works and how to prevent mis-trigger of this function will be explained together.



Summary of

How to prevent mis-trigger in these modeless functions

Not entering mode means those function will be turned on all the time, the advantage of this is obvious, it is much simpler, considering our natural gestures in the communication are dynamic and random, the vital part of these modeless functions is preventing mis-trigger, after all, nobody wants random things happens on their screen during their meeting.



Two of three

Together there are three conditions, gesture of left hand, gesture of right hand and the positional relationship between these two hands. Any modeless function should at least meet two of the three conditions.

Why those functions

Why not making all functions of the concept in this zone? There is the possibility of making those functions to be modeless because of the research of our natural gestures. From the workshop I mentioned before, our natural way to represent those shapes, lists requires two hands, which leaves the possibility to prevent mis-trigger problem, but some effects are more natural to achieve with just one hand, then we need to enter some mode.

Refined sketch

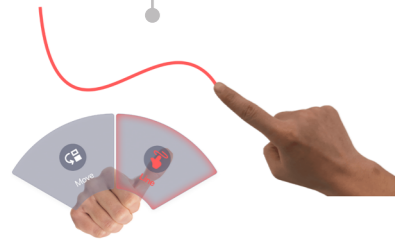
Function with mode

- Draw lines
- Grab & move elements
- Summary -mode

Functions with mode-Draw lines

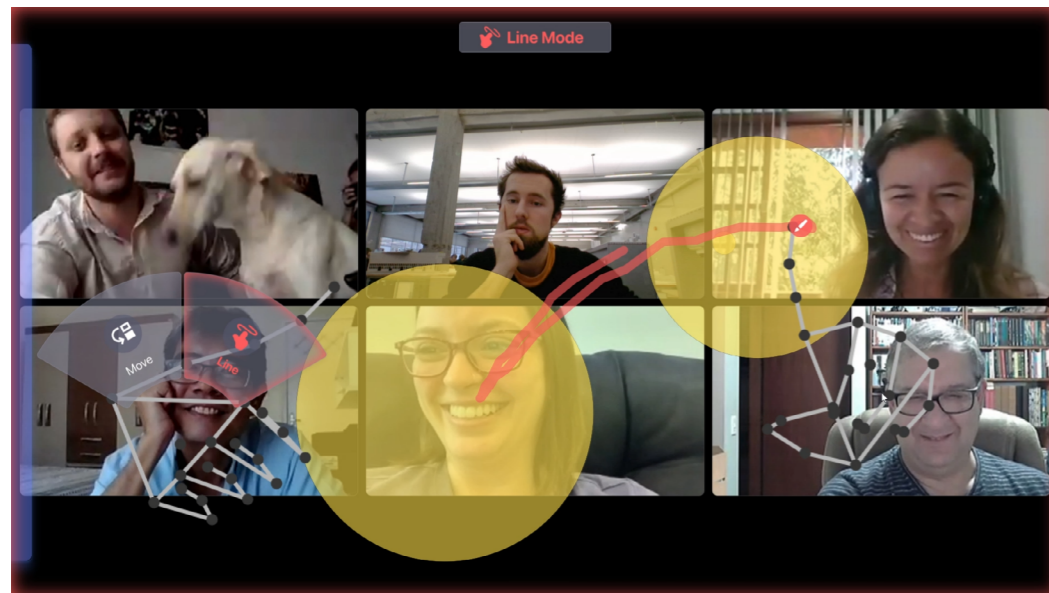
User needs to perform the certain gesture to enter the mode, that means the function needs to be turned on and then it can be triggered.

Used component



Draw lines

Use one hand to perform the setting gesture and use the thumb as the pointer to choose line mode, then use the other hand perform the component and start to draw.



<https://ahocloud.box.com/s/eszjctczxqde8oqkw56h8rweubuq3gx6>

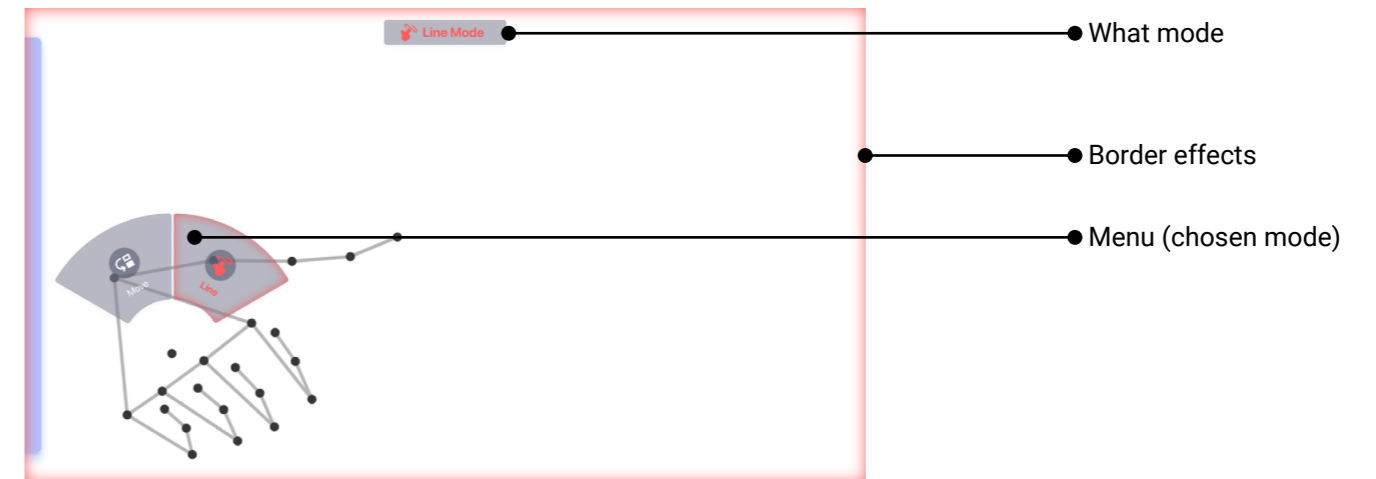
The concept of drawing lines is not meant to create some delicate illustration, it serves to help the user create more flexible shapes or explain the relationship between elements.

How it works and prevent mis-trigger

When user perform both gestures and entering the mode, the machine will calculate the coordinates of the index fingertip and use that to draw. Obviously it avoids mis-trigger by entering the mode.

Indicate what mode it is now

Not only the menu on speaker's hand will indicate what mode user is in now, but there will also be the effects below showed to the user, and make it more clear for him to know the status.



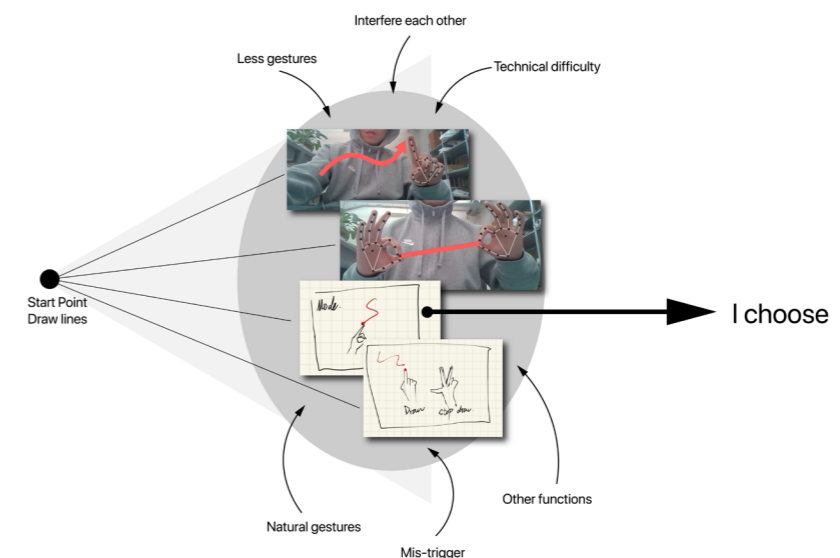
Indicate where the effect occurs

When the drawing lines function has been triggered, there will be an indicator inform user which fingertip will cause the effects.



More possibility

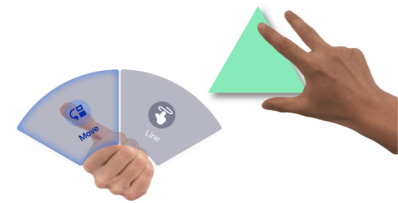
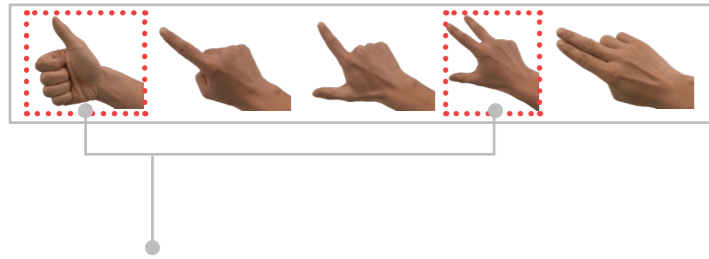
There is more possibility about drawing a line, for example, instead of line, it can be an arrow, straight line, create a line by confirm two points, etc. With the limited time, I cannot develop every of my sketch and test/evaluate, here I choose this way based on my principle (it will be mentioned later) and analysis, but it doesn't mean the only right solution, there is definitely more possibilities.



Functions with mode-Grab & move elements

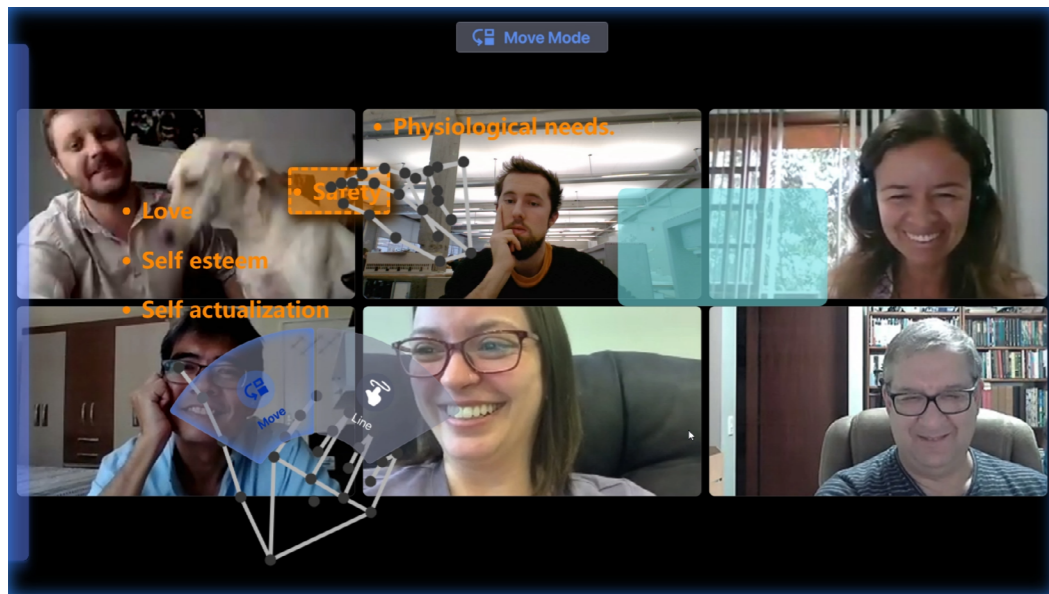
Another mode in this concept is the moving mode. In the moving mode, user can use gesture to grab and move those shapes or words and it doesn't matter who created those elements.

Used component



Grab the elements

Use one hand to perform setting gesture and another hand perform the certain component, use that same hand to grab elements on the screen.



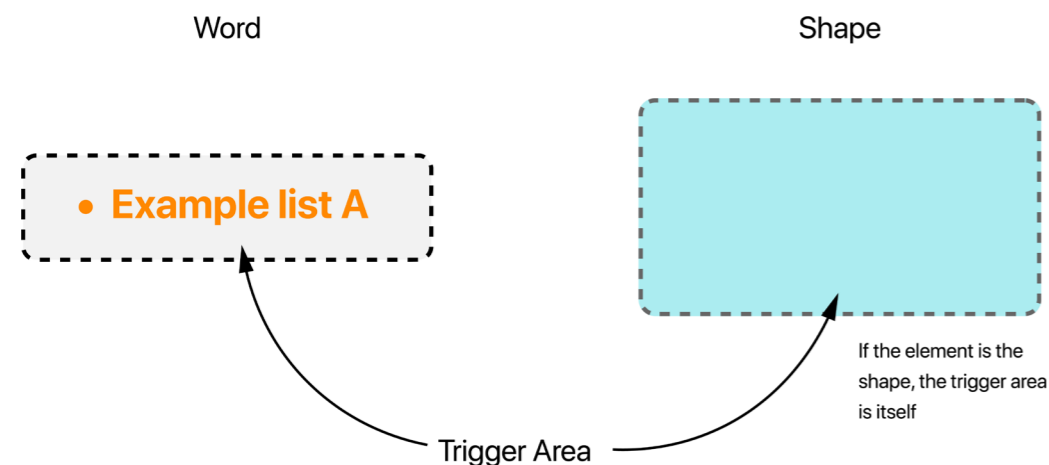
<https://ahocloud.box.com/s/6jxi7hkxf38zupuokb2opz90sz4zdebv>



Please scan the QR code or click the hyperlink under the image to see the real effects

Trigger area of elements

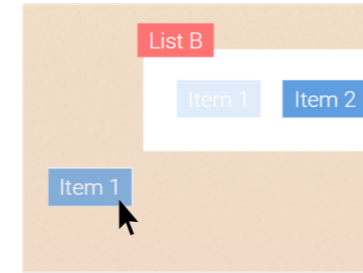
In order to manipulate the elements user created, there should be a trigger area in each element.



Difference between move object with mouse and hand

How we drag now

People gets used to drag elements with mouse, if there is a touchable device, with fingers.



Drag with mouse

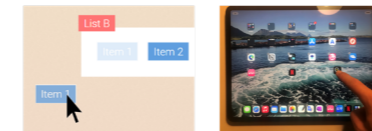


Drag with finger

Drag with gesture

Unlike mouse this classic interactive input device, there is no basic research of how we can drag with gesture, moreover, I would view drag with gesture is more natural while drag and move with mouse/finger is relatively more abstract, because dragging with gesture reflects how we grab things in physical life.

Drag with mouse



Drag with finger



Drag with gesture



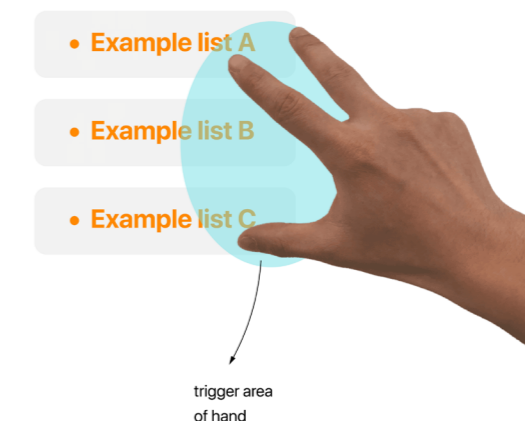
Abstract

Natural

Therefore, I compared gesture dragging and mouse dragging to see what problem there may be.

Trigger area - Precision

With mouse, we always have a cursor to inform us where is my current position, it's very precise and the trigger area is basically a point. But with hand, if we want to grab something with the grabbing gesture, the trigger area is big and dynamic. Imagine if there are several elements together on the screen, when using gesture, how the machine knows which one user is on top of, furthermore, if the machine doesn't know that, how can it give user feedback.



Difference between move object with mouse and hand

How to trigger

Taking dragging element with mouse as an example, user can click and drag, here click means trigger the function, and then user can move. With gesture, there is no such operation, how to trigger the dragging function with gesture?



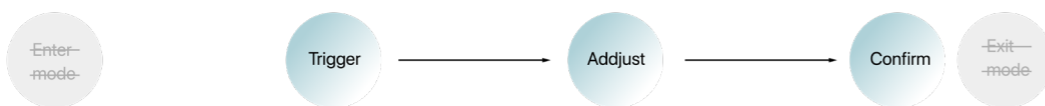
How to confirm

Furthermore, how to confirm the effects, when using mouse to drag element, we can release to end the moving and confirm the new place, but how to do it with gesture?

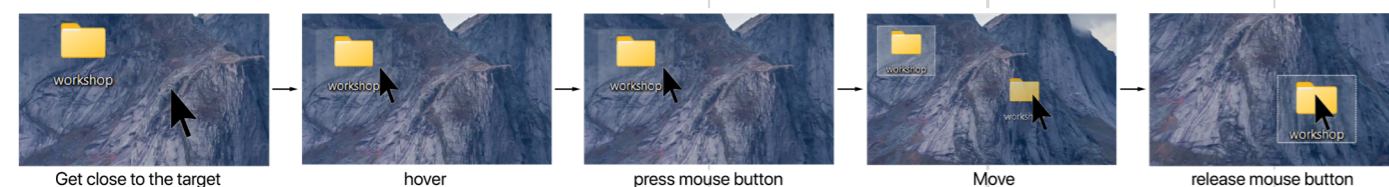
Observation of physical world

Since the grabbing gesture reflects people's grabbing action in the physical world, the observation and analysis of how people grab and move object in physical world would be helpful. After that, a comparison of grabbing with hand, dragging with mouse and our five stages theory was conducted to find the pattern.

5 stages



Mouse



Hand

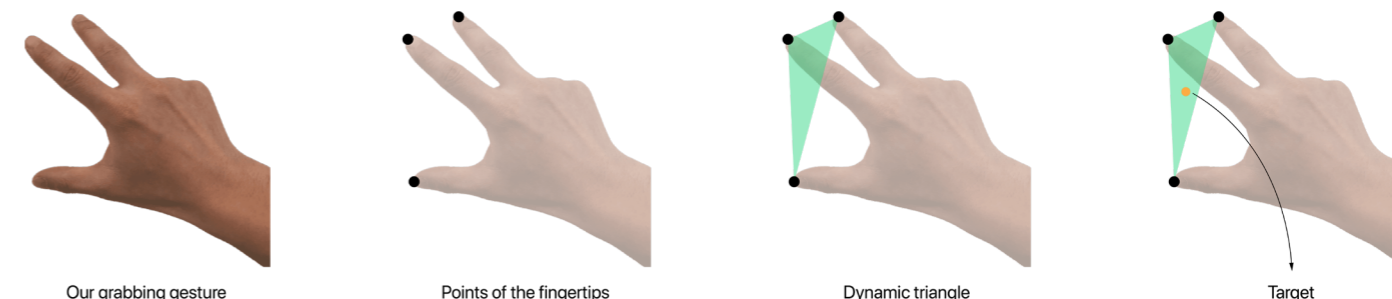


My solution

After those research, comparison and analysis, I come up with this potential solution.

Trigger area - dynamic triangle

With the grabbing gesture we can get the coordinates of index fingertip, middle fingertip and thumb fingertip, by these three points, a triangle can be drawn and middle point of the triangle is the target/ trigger area.

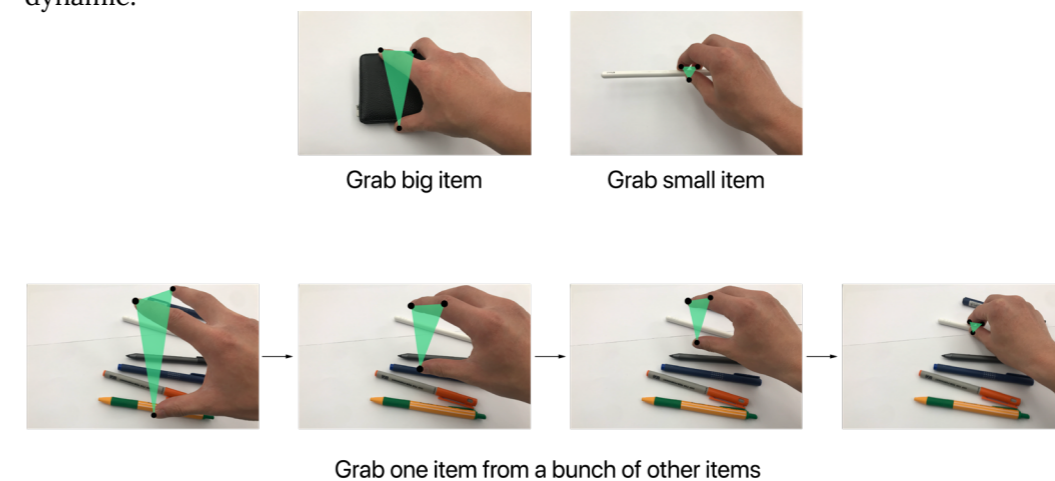


Precision

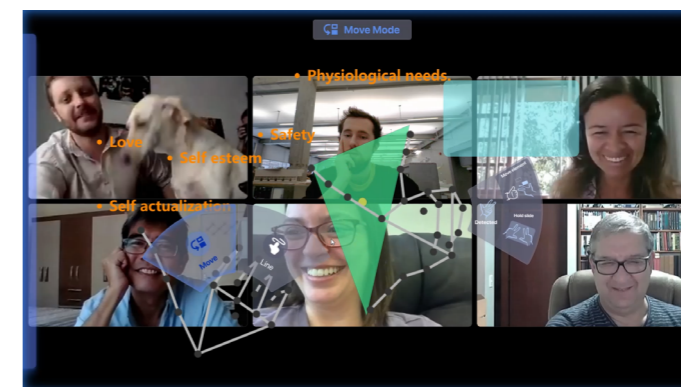
Through this way, we can transfer the big trigger area into a small point, so the machine can be sure about which element user is on top of, hence give the feedback (e.g., hover effects of the hyperlink)

Dynamic

Since the final trigger area is a point, why not just choose one point from the hand? Our grabbing action is not a static image, when people grab the big item, they will open the hand bigger, when it is the small item, they will open the hand smaller, and if they want to choose one single small item from a bunch of other items, the angle of hand opening will change from large to small, that shows our target area is dynamic.



By calculating the middle point of this dynamic triangle, we can simulate the natural grabbing to the greatest extent.



Please scan the QR code or click the hyperlink under the image to see the real effects

<https://ahocloud.box.com/s/v1yjjd0q5w8vlvit39w39ai1tyojfe6>

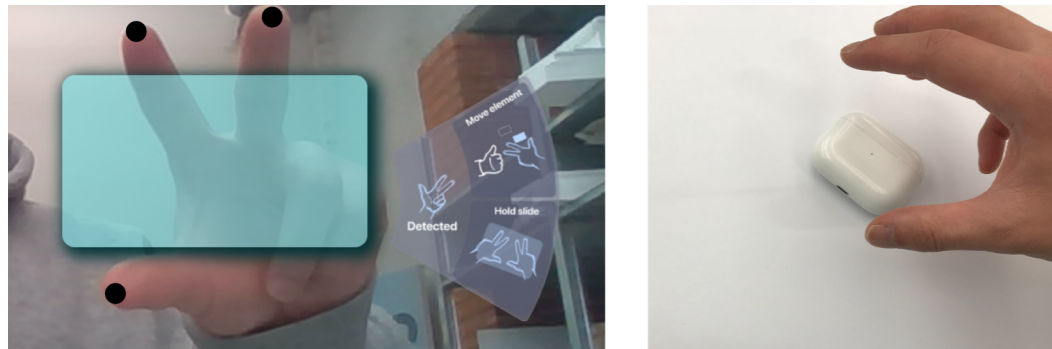
How to trigger

When all three fingertips are inside the trigger area of the element, then the moving function will be triggered and user can move the object. It is same as we start to move object in the physical world.



How to confirm

Similar with triggering the function, when all three fingers are outside the trigger area of the element, then the moving will be ended and the element will be in the new place. Also same as we drop object in the physical world.



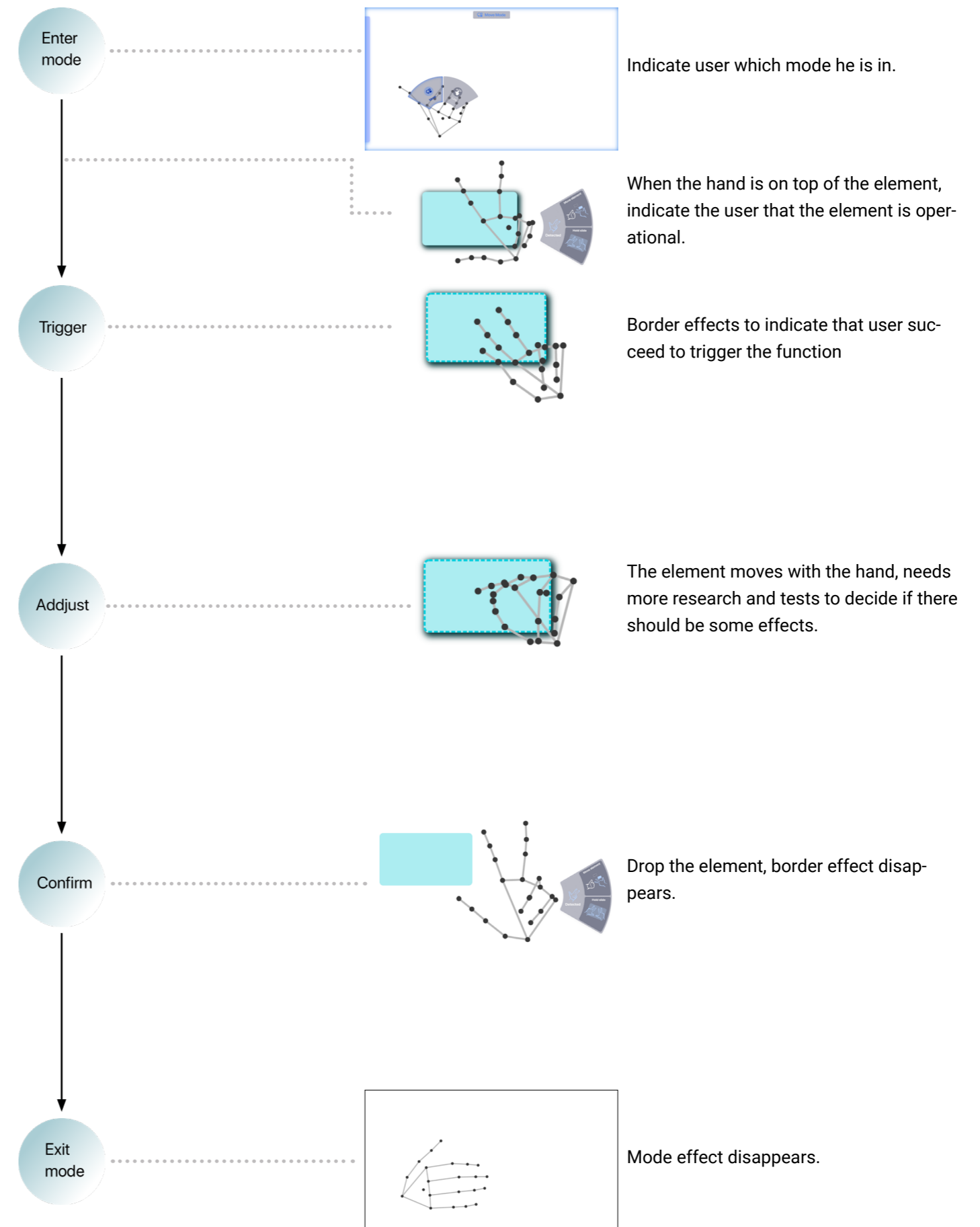
<https://ahocloud.box.com/s/gdhg20bynoyldslqt3bunt3vgjkm5c>



Please scan the QR code or click the hyperlink under the image to see the real effects

Indicate status by stage

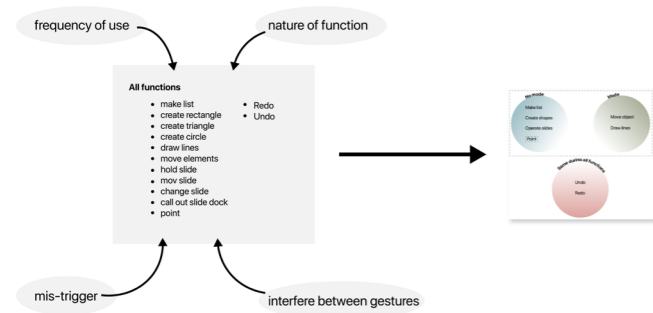
As mentioned before, when designing those functions, there are 5 stages we need to think about, in addition, the feedback of each stage needs to be considered, taking this move/grab function as an example



Summary of the mode

Why those functions-natural gestures

As mentioned before, there are many factors to be considered when decide which function needs the mode which does not. But the main aspect which I considered is using natural gestures



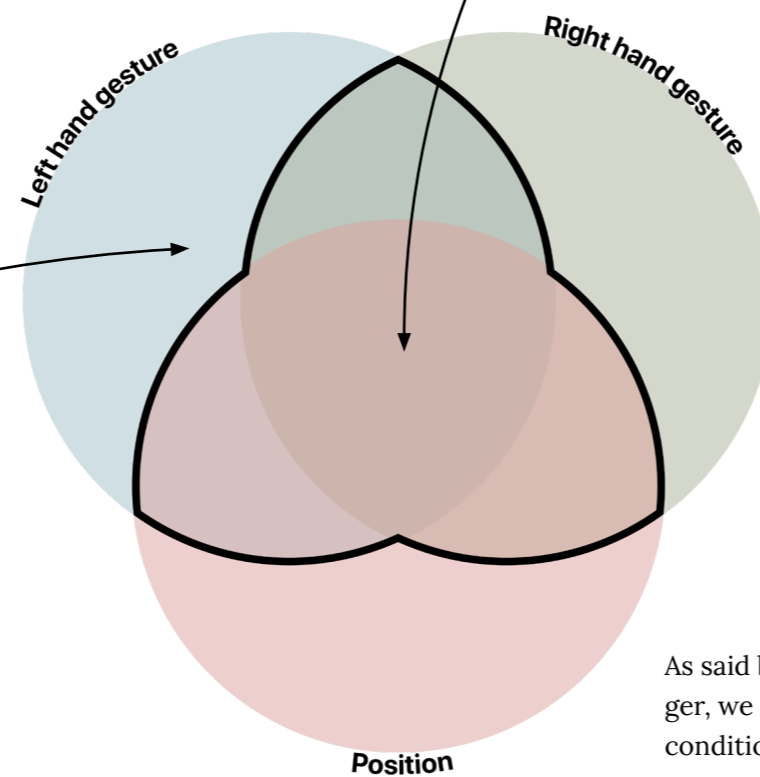
Advantage of using natural gestures

- In line with people's intuition
- Easy to remember for regular user
- Not awkward under the meeting context

Single hand gesture



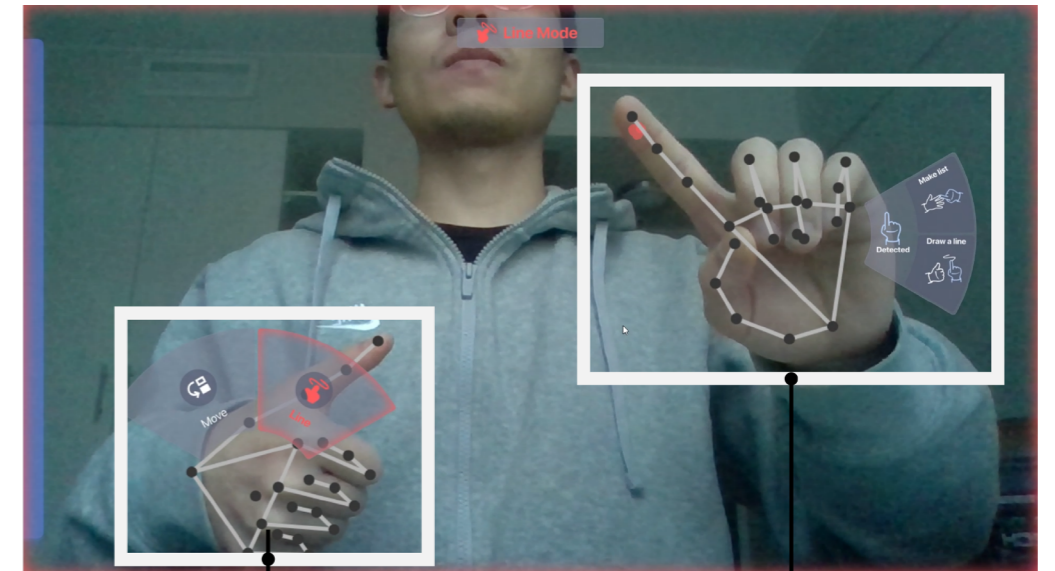
From the workshop I did, images above are the most common gesture of drawing lines and grabbing something, it is a single hand gesture and contain very broad meaning as well (e.g., drawing line gesture can also mean first, pointing at something), that means those gesture will be used frequently in the conversation and leads to mis-trigger problem in this case.



As said before, to prevent mis-trigger, we need meet at least 2 of 3 conditions

How the mode works

By entering the mode with one hand, there is the possibility to use the natural gesture without worrying about mis-trigger issue.



Mode / setting gesture

Essentially, this setting gesture serves to prevent mis-trigger by adding another condition (the appearance of this gesture).

Natural gesture

With the limitation of the setting gesture, the trigger gesture can be as close as possible to people's natural gesture.

Exit the mode

Unlike our normal understanding of "entering" the mode, here we enter the mode by just showing the setting gesture, so whenever user wants to exit mode, they can just stop the gesture.

Principles

Here are the principles behind these refined sketches, just like sketches, the principles are not the only right either. Some of the principles have been partly mentioned in the refined sketches phase, here I will go through all of them together.

Principles behind them

- Talk to people rather than machine
- Extract natural gestures
- Less gestures
- Prevent mis-trigger
- Indicate on the hand
- Don't break flow

Talk to people rather than machine

I feel like you are talking to the machine rather than talking to me



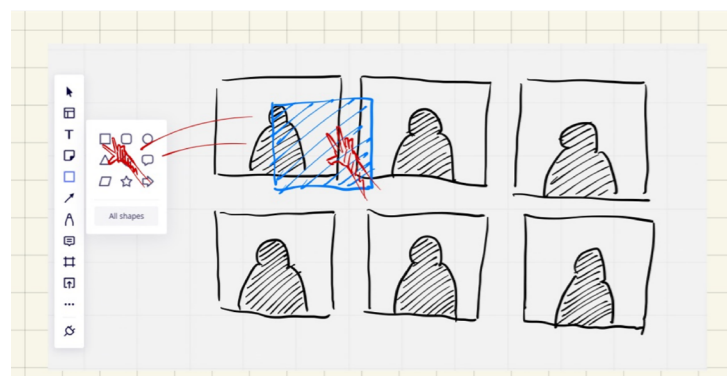
One of the remote testing

During the exploration of the concepts, not only I draw lots of sketches, but I also made the functional prototype to get inspiration and test with people. In one of the tests, the candidate told me that he felt like I was talking to the machine rather than him.

That comment helped me reflect about my topic and leads to my first principle: Talk to people rather than machine. As the research I mentioned in the beginning, our natural gestures are extremely important in the communication, so I think even with gesture recognition this technology, our gesture should not just be another input source under the context of meeting, it should serve both as the visual information itself and also the input source. Admittedly, under the framework of existing technology, it is extremely hard to achieve both aspects at the same time, but I try my best to make the trigger gesture meaningful for both speaker and audience.

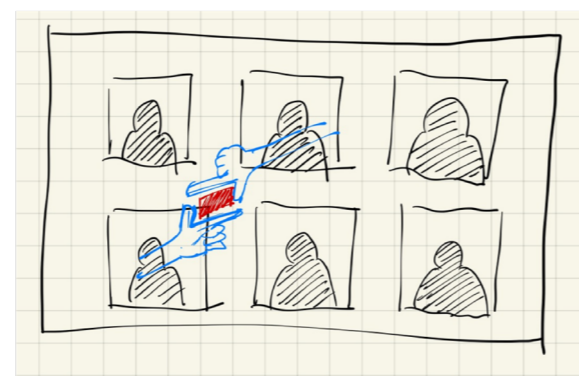
To explain this principle better, I will take two of my sketches as example, I would choose sketch B because in sketch A, dragging a shape to the screen from the menu, this action is talking to the machine rather than the audiences, if we take away this menu and shape, the whole the action will be meaningless to the audience. On the contrary, in sketch B, even there is no shape created, the gesture still represents the shape, and it is meaningful to the audiences.

Sketch A



There is a shape menu on the side, speaker can grab the shape to the transparent "blackboard."

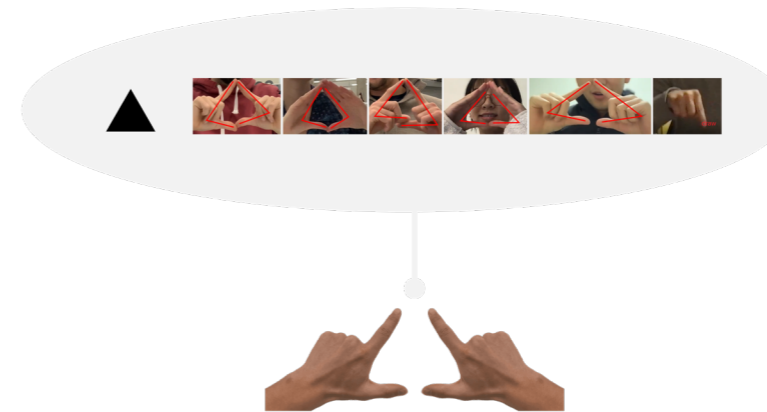
Sketch B



Use gesture to create the shape directly.

Extract natural gestures

Obviously recognizing natural gestures and achieving functions with them would be the best solution, but those natural gestures are dynamic, they are also slightly different according to different people, even for one same person, he would use different gesture to represent the same idea in one communication, furthermore, once we set bonding between the function and the gesture, then the gesture is not natural anymore. So I think we should first gather the gestures with the same meaning of the function, then put them together and find the pattern, extract features from those natural gestures, in this way we can find the gesture that's closest to our natural one, suits most people and technically, easy to be detected by machine.

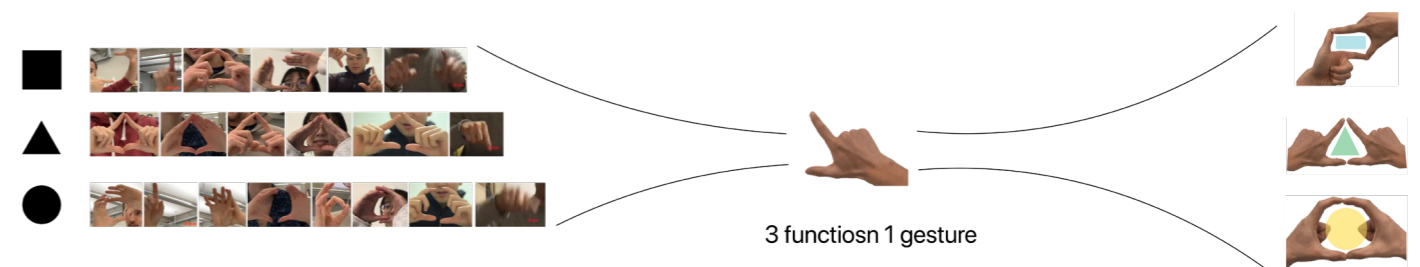


Example: put most people's natural gesture of representing triangle together and we can find that the index and thumb is the key.

Less gestures

For regular user, if there are multiple functions, I think we should not simply use one gesture to correspond one function, even the gesture is close enough to our natural gesture. Taking my core concept as an example, if there is an individual gesture for each function, together there will be about ten gestures, and that is hard to remember, not mention that some gestures may be very similar, make it even harder to memorize.

My approach to limit the gestures: Extracting the features from the natural gestures and using them as the components, by combining these components, we can trigger different functions. Moreover, according to this method, it's also easier for the user to remember how to use the function, because we can tell them about the pattern behind those gestures, rather than just showing them static gesture pictures.



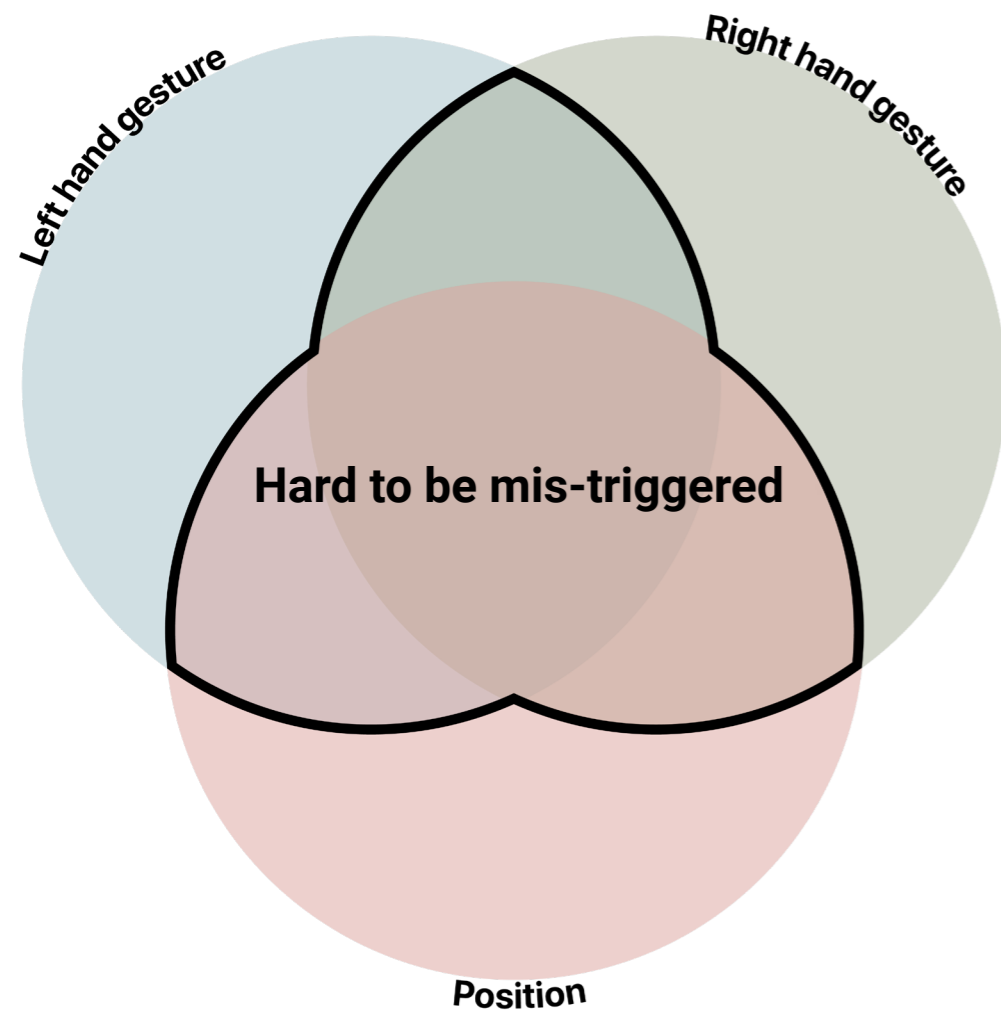
3 functions 1 gesture

Principles behind them

Prevent mis-trigger

During our communication, there are lots of natural gestures and hand movements which are hard to predict. This feature makes it different with applying gesture recognition in the regular situation such as switching music on mobile phone. Because of this unpredictable hand movements, preventing mis-trigger problem becomes the primary issue, I think nobody wants random things happen on their screen during their talking in the meeting.

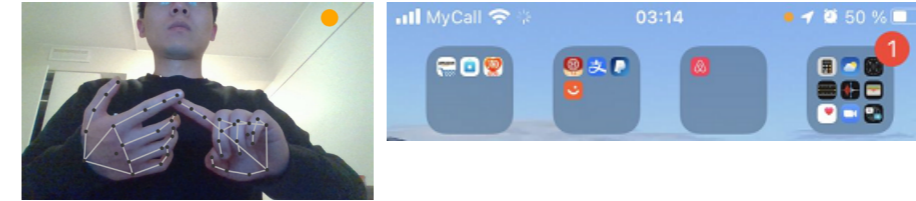
As I mentioned in those refined sketches, my method of preventing mis-trigger is setting left-hand gesture, right-hand gesture and position these three conditions. In order to trigger the function, that user's hands should meet at least two of three conditions. The reason behind it is simple the probability of showing the certain gesture with one hand is high, but two hands with the certain gestures is lower, two hands with certain gestures and special positional relationship is even lower.



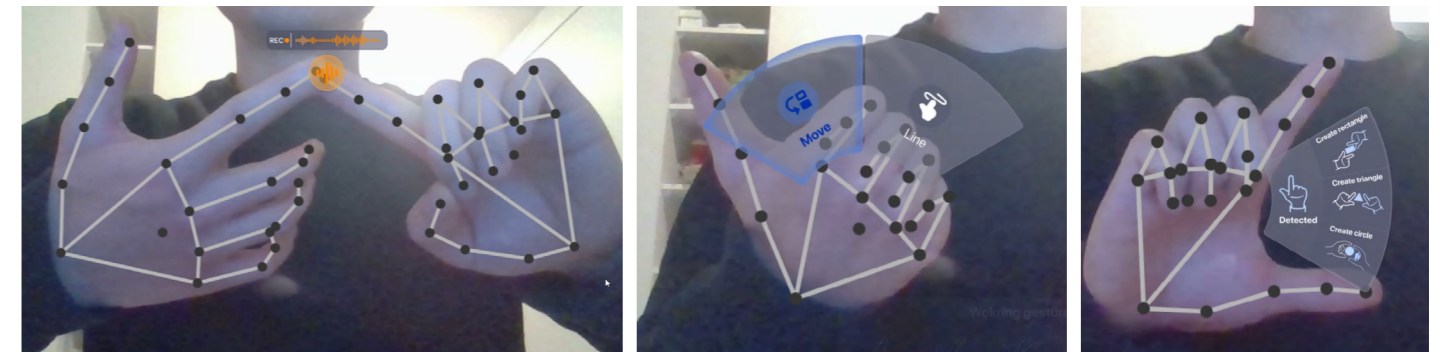
Indicate on the hand

This principle also comes from a testing with other people, there is one prototype that every time speaker making a list with gesture, there will be a dot showing on the corner of the screen to indicate the user that he is being recorded, this effect is just like the effects when we record with cell phone. But from the testing, people focused their attention on the hands rather than the corner of the screen, and it makes complete sense, because people are creating things with their hands. This feedback leads to my final principle, all indication should happen on the place triggering it, which is our hand.

Early prototype - indicate in corner



Indicate on hand



Don't break the flow

This principle is quite subjective, when I thought about those ideas, I avoided asking user to use mouse. Because I think using mouse will break the flow of people's speech, it will also influence speaker's natural gestures. But that's just my principle behind those ideas.

Test and discuss 2.0

After I finished the refined sketch, I conducted another round of testing and discussion. Mainly with the experts in the meeting field.

Test and discuss 2.0

- Test and discuss with experts
- Test and discuss with regular people

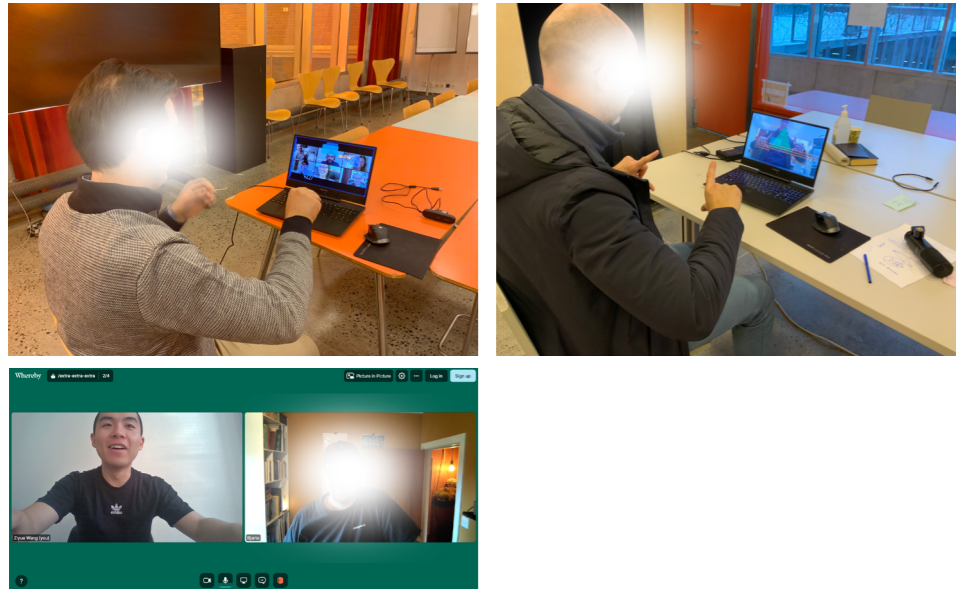
Test and discuss 2.0

Test and discuss with experts

After coming up with the refined sketches, I tested them with the experts in remote meeting field.

Who I tested and discussed with

- Agustin Vargas, product manager from Cisco
- Xxx, xxxx from Cisco
- Fredrik Matheson, Creative Director from Bekk
- Bjarte Misund, former designer of Whereby



Some interesting feedback and reflections

- **0~1**
How to make user, who doesn't know anything about those functions or even gesture interaction, start to use those functions. How to teach them? An onboarding stage or give them cues during their trying?
- **Cues**
There should be some cues to tell user that this screen is different, they can perform some gestures towards it. Furthermore, when user are in the middle of triggering the function, there should be some cues to guide user achieve the function
- **Not showing whole virtual hands**
Those dots of the virtual hands can be very distracted, maybe those dots are not necessary to the user, furthermore, if some functions are just triggered by several fingers, maybe we can just show those fingers rather than whole hands.
- **Examples of where to use those functions**
The usage scenarios of some functions are quite obvious, for example making a list, but some functions are hard for users to know where they can use them, e.g., creating shapes, so maybe there should be some examples that can inspire users.

Test and discuss with other user

I also tested the prototype with regular users



Some interesting feedback and reflections

- **Learning curve**
Most people are not familiar with gesture interaction, there will definitely be a learning curve of using those functions. Moreover, how to make user be willing to learn to use those functions? Maybe some gesture interaction games that help user to learn/get used to those functions?
- **Hard to say**
As a very unique function, there is no similar experience that can be used to compare with, hence it's hard to tell how it feels like when really using those functions in the meeting.

Summary & Reflections

Details

In this project, I paid most of my attention to explore how to use gesture interaction in different way to support explaining ideas, there are a lot of details I don't have time to cover, for example what kind of UI elements would be better, what color would be most recognizable in different speaker's messy background, how to adjust the size of those UI elements according to the size of the hands etc. I believe there will be a lot of research we can do around those details.

More aspects

Onboarding function, how to guide user when they perform the gesture in the wrong way, how will those gestures and functions be different under different culture, what will happen when several people speak together, etc. There are lots of different aspects that we can dig into, but started with this new technology, I need to dive into one certain aspect to see the possibility first, then I can step back and think about those other aspects.

More possibility

As I mentioned in many places of the report, the purpose of this project is not giving the only perfect solution, it's about showing the possibility, although I integrate some of those refined sketches based on my principle, but each single function in that group can be a start point and extend more possibilities. For example, in that pointing function, not just presenter's video going there, it can also zoom in the place he points. When both presenter and audience are moving their video through pointing function and their video bumps together, something can happen there.

Foundation

Although the project is about showing the possibilities, but there are other aspects that I think it's also valuable.

When designing around gesture recognition, we can't just view that the machine can understand our gestures and start imagination from there. In fact, by using machine leaning to recognize people's hands through the camera, we can have the coordinates of our hand joins map on the screen. Those coordinates are all what we have, so exploring how can human interact with machine/computer through those hands coordinates is an important part of my project. If we don't figure out those basic HCI problems, the further imagination would be unrealistic. I think my thoughts of those basic issues such as how to trigger the function, how to confirm effects, how to prevent mis-trigger etc., can also be valuable, it can be a foundation for others to go further.

References

<https://www.youtube.com/watch?v=p2J7wSuFRl8>
<https://atap.google.com/soli/>
<https://www.microsoft.com/en-us/research/project/project-prague/>
<https://www.ultraleap.com/>
<https://www.toolbox.com/collaboration/remote-support/guest-article/5-things-that-are-making-your-virtual-meetings-ineffective/>
<https://www.boredpanda.com/funny-video-meetings-conference-call-fails-quarantine/>
<https://demodesk.com/blog/online-meetings/most-common-issues-explained><https://miro.com/blog/remote-meeting-worst-practices/>
<https://www.prdaily.com/why-we-should-stop-pretending-virtual-meetings-are-working-and-how-to-fix-them/>
<https://www.youtube.com/watch?v=tKEtm3HCrsW>
https://www.youtube.com/watch?v=iP8J35dZ_W4
https://www.youtube.com/watch?v=oMcqz_O_dPU&list=RDLV53yPfrqbpkE&index=4
<https://www.youtube.com/watch?v=JpTDF4w-6Yc&list=RDLV53yPfrqbpkE&index=5>
<https://www.youtube.com/watch?v=BPJ5C7G0yVk&list=RDLV53yPfrqbpkE&index=7>
<https://www.youtube.com/watch?v=vCiLMLC2Rhs&list=RDLV53yPfrqbpkE&index=15>
<https://www.youtube.com/watch?v=jma9hFQSCDk>
<https://www.youtube.com/watch?v=h4ROqE4SMYA>
<https://www.ideo.com/blog/why-gesture-is-the-next-big-thing-in-design>
<https://www.masterfile.com/image/en/6109-06196479/students-chatting-during-class>
<https://www.gettingsmart.com/2015/10/26/talking-math-6-strategies-for-getting-students-to-engage-in-mathematical-discourse/>
<https://lifeasahuman.com/2015/arts-culture/science/andy-beal-the-numbers-man-who-inspired-a-generation/>
<https://www.youtube.com/watch?v=vtIzMaLkCaM>
https://www.nomodes.com/Larry_Tesler_Consulting/CV.html
<https://journals.openedition.org/anglophonia/1200?lang=fr>
<https://www.cssscript.com/demo/lightweight-js-sorting-library-with-native-html5-drag-and-drop-sortable/>
<https://miro.com/>
<https://www.youtube.com/watch?v=4-RZRLdBpFc>
<https://interconnected.org/home/2020/06/04/projectors>
<https://mediapipe.dev/>
<https://handsfree.js.org/>
<https://lodash.com/>
<https://dmitrybaranovskiy.github.io/raphael/>