

Towards Accessible Conversations in a Mobile Context for People who are Deaf or Hard of Hearing




We're going down the stairs now...

Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich
University of Washington, Seattle
Gallaudet University

A grayscale photograph of a group of people sitting around a large table in a meeting or lecture setting. The people are engaged in conversation and looking at documents or laptops on the table. The scene is dimly lit, with the text overlaid in the center. The text reads: "Prior work have investigated communication challenges of DHH people in **stationary contexts** such as group meetings and lectures." The words "stationary contexts" are highlighted in yellow.

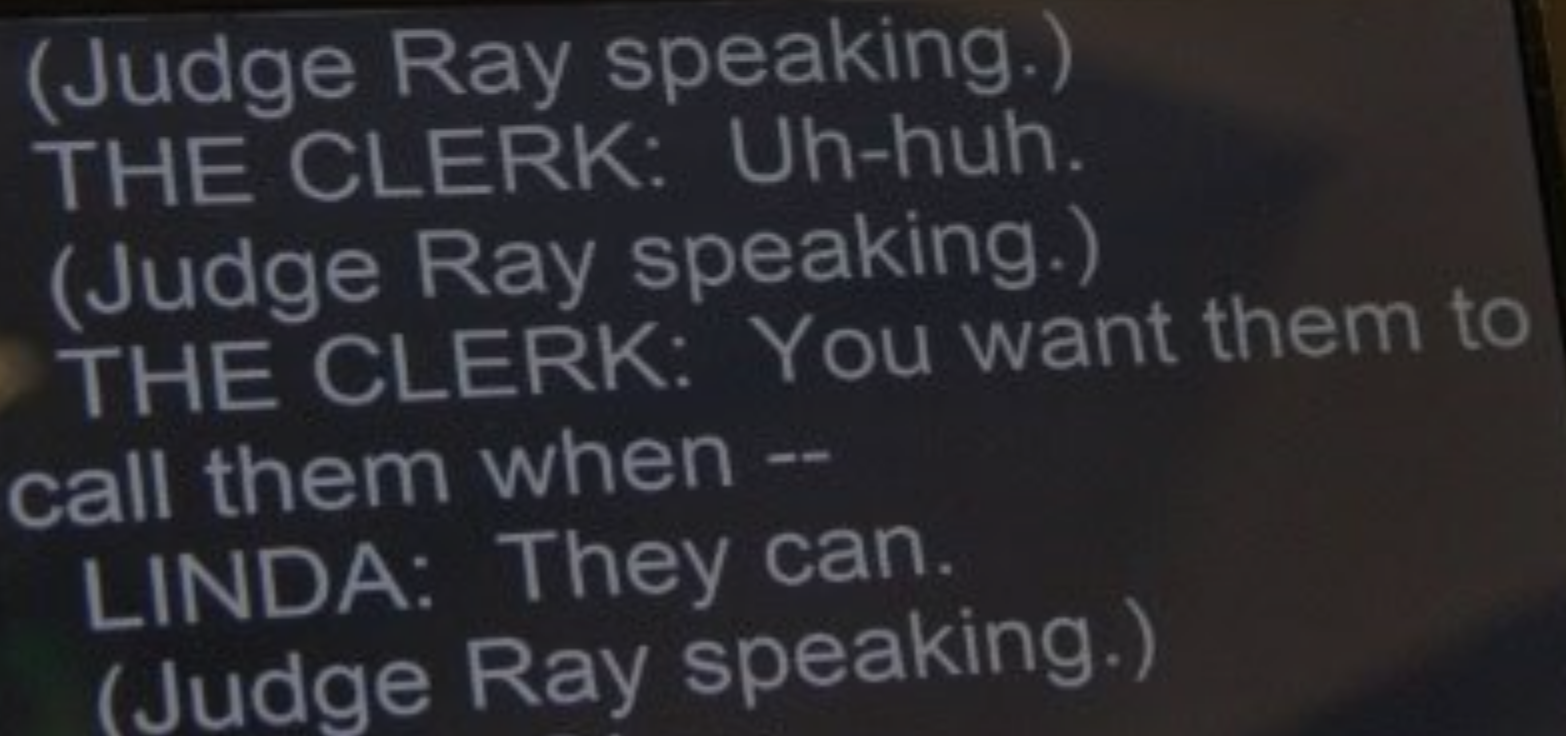
Prior work have investigated communication challenges of DHH people in **stationary contexts** such as group meetings and lectures.

A photograph of four business professionals walking on a city street. From left to right: a Black man in a dark suit and blue tie, a woman in a black blazer and dark pants, a man in a light blue blazer and light blue shirt, and a man in a dark suit and dark tie. They are all smiling and appear to be in conversation. The background shows a city street with buildings, trees, and a traffic light. The image is dimmed to allow text to be overlaid.

Prior work have investigated communication challenges of DHH people in **stationary contexts** such as group meetings and lectures.

Moving conversations (e.g., walking) could present **new challenges** such as varying background noise and **needing to balance visual attention** between looking at the speakers and looking ahead.

Moreover, assistive technologies like real-time captioning have been traditionally designed for stationary context and are **not conducive to mobile scenarios.**

A close-up photograph of a screen, likely a tablet or smartphone, displaying a transcript of a courtroom scene. The text is white on a dark background and is slightly blurred, suggesting the screen is being viewed from a distance or is in motion. The transcript includes dialogue between Judge Ray, a Clerk, and Linda.

(Judge Ray speaking.)
THE CLERK: Uh-huh.
(Judge Ray speaking.)
THE CLERK: You want them to
call them when --
LINDA: They can.
(Judge Ray speaking.)

AIMS



A group of business professionals in a modern office hallway. A man in a dark suit and tie is in the foreground, looking towards a woman in a black blazer and white shirt who is smiling. Other people are visible in the background, including a woman in a black dress holding a blue folder. The scene is dimly lit, with a dark overlay on the image.

AIMS

- 1 To investigate the **needs of DHH people in mobile conversations** such as walking and transit.

AIMS

- 1 To investigate the **needs of DHH people in mobile conversations** such as walking and transit.
- 2 To study the possibility of **captions on head mounted displays (HMDs)** to support those needs.

THIS PAPER





THIS PAPER

- 1. Study 1:** Formative interview with 12 DHH participants on challenges, communication strategies, and future captioning technology.

THIS PAPER

- 1. Study 1:** Formative interview with 12 DHH participants on challenges, communication strategies, and future captioning technology.
- 2. Study 2:** Evaluation of a proof-of-concept HMD-captioning prototype with 10 DHH participants in a walking scenario.

OUTLINE

Background
and Past Work

OUTLINE

Background
and Past Work



Study 1:
Interview

OUTLINE

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype

OUTLINE

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

AUTOMATIC SPEECH RECOGNITION (ASR)



Captions can be generated in two ways:

AUTOMATIC SPEECH RECOGNITION (ASR)



TRAINED TRANSCRIBER



TRAINED TRANSCRIBER

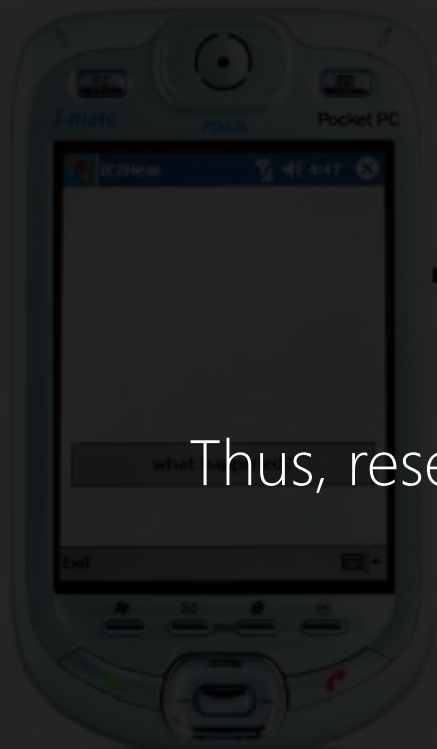
We used a trained transcriber (or real-time captioning).

TRAINED TRANSCRIBER

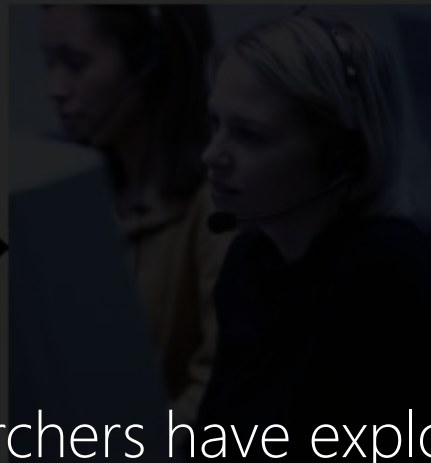
A person wearing a white lab coat is seated at a desk, typing on a white keyboard. The desk is cluttered with various items, including a laptop displaying a blue screen, a purple water bottle, and several papers. The background shows a framed picture and a red apple. The overall scene is dimly lit, suggesting an indoor office or clinical environment.

We used a trained transcriber (or real-time captioning).

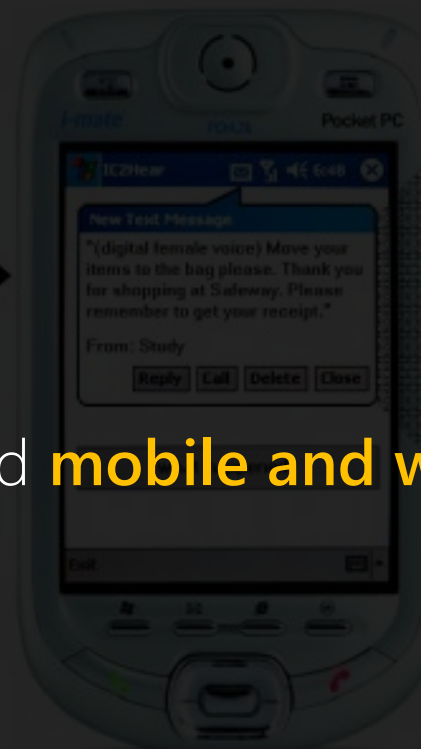
Captions from a trained transcriber are typically shown on a **laptop or a large shared screen.**



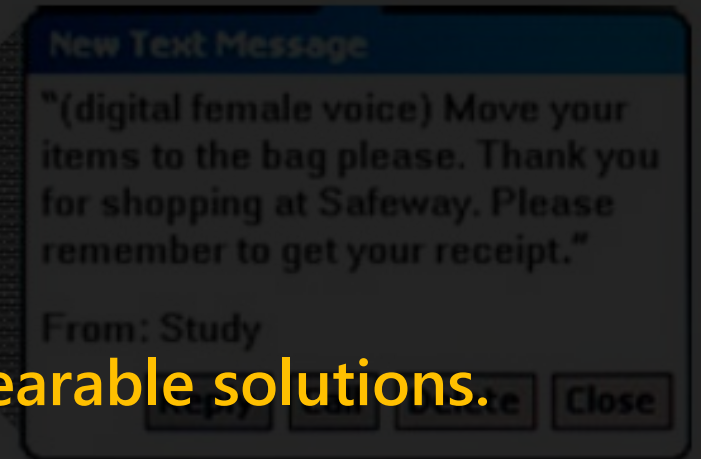
(1) user needs sound info & clicks "what happened?" button



(2) the past 30 seconds of audio is sent to translator



(4) user gets text message describing sound



Thus, researchers have explored **mobile and wearable solutions.**



(1) user needs sound info & clicks "what happened?" button

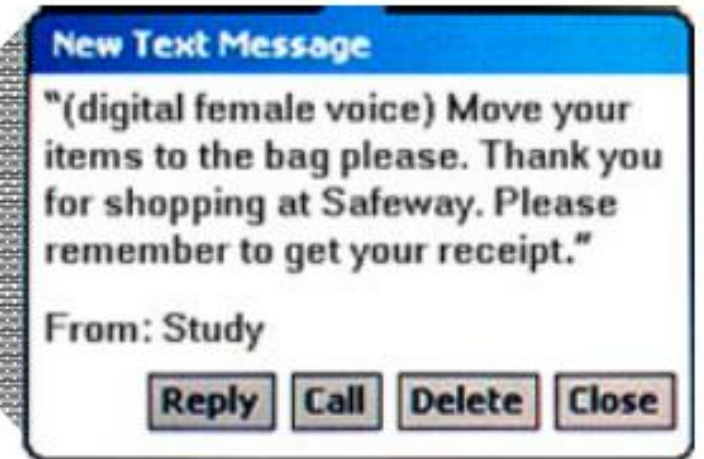


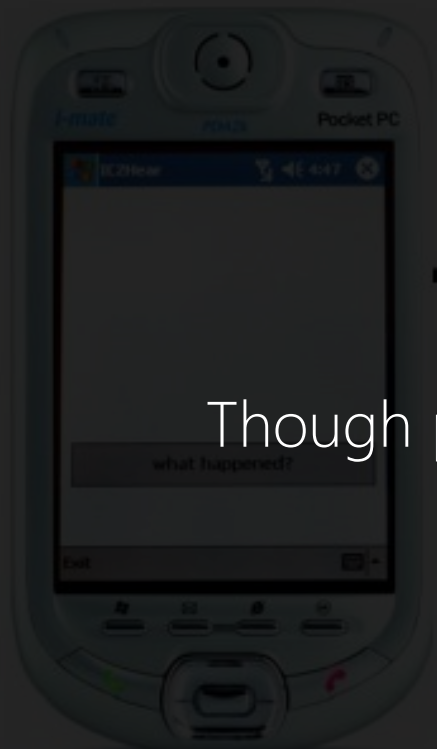
(2) the past 30 seconds of audio is sent to translator

(3) translator translates audio & sends to user in a text message

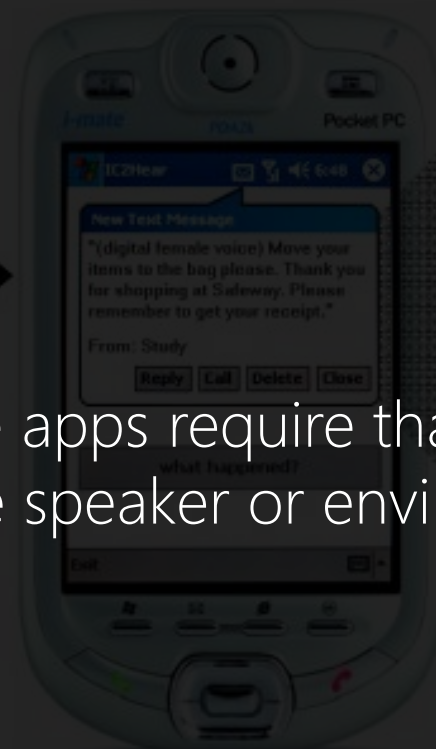


(4) user gets text message describing sound

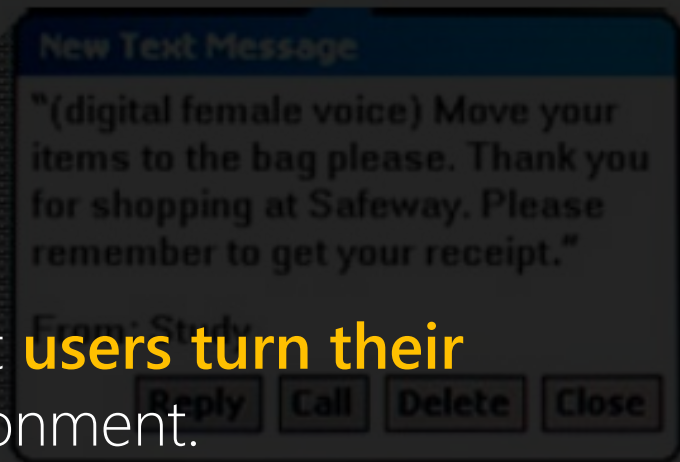




(1) user needs sound info & clicks "what happened?" button

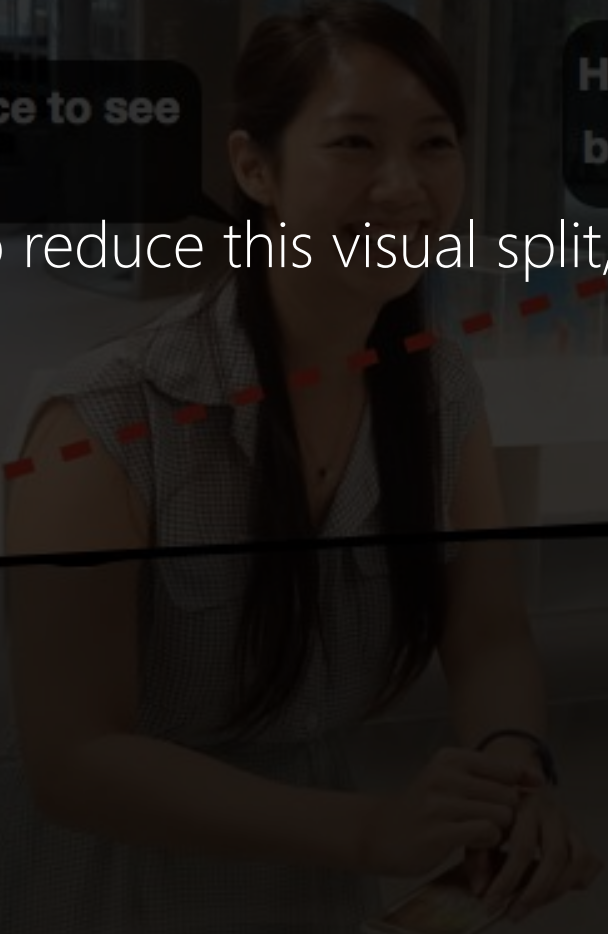


(4) user gets text message describing sound

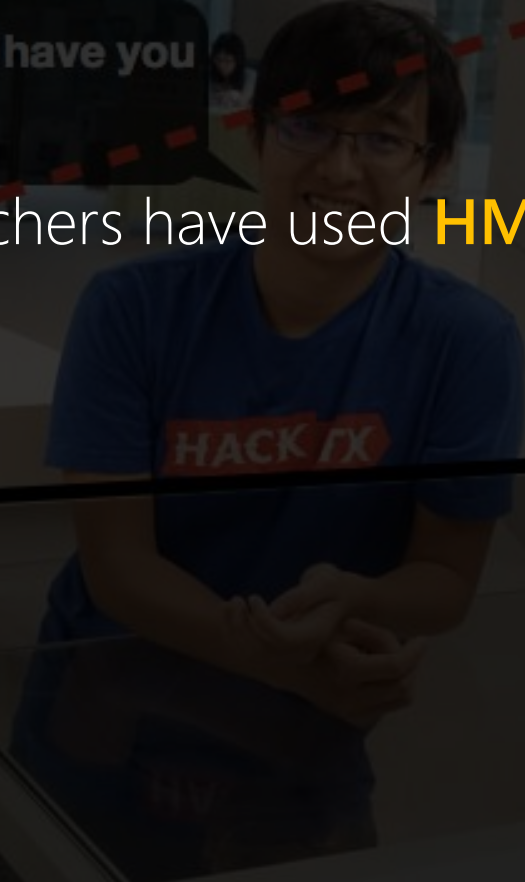


Though portable, smartphone apps require that **users turn their gaze away** from the speaker or environment.

(2) the past 30 seconds of audio is sent to translator
(3) translator translates audio & sends to user in a text message

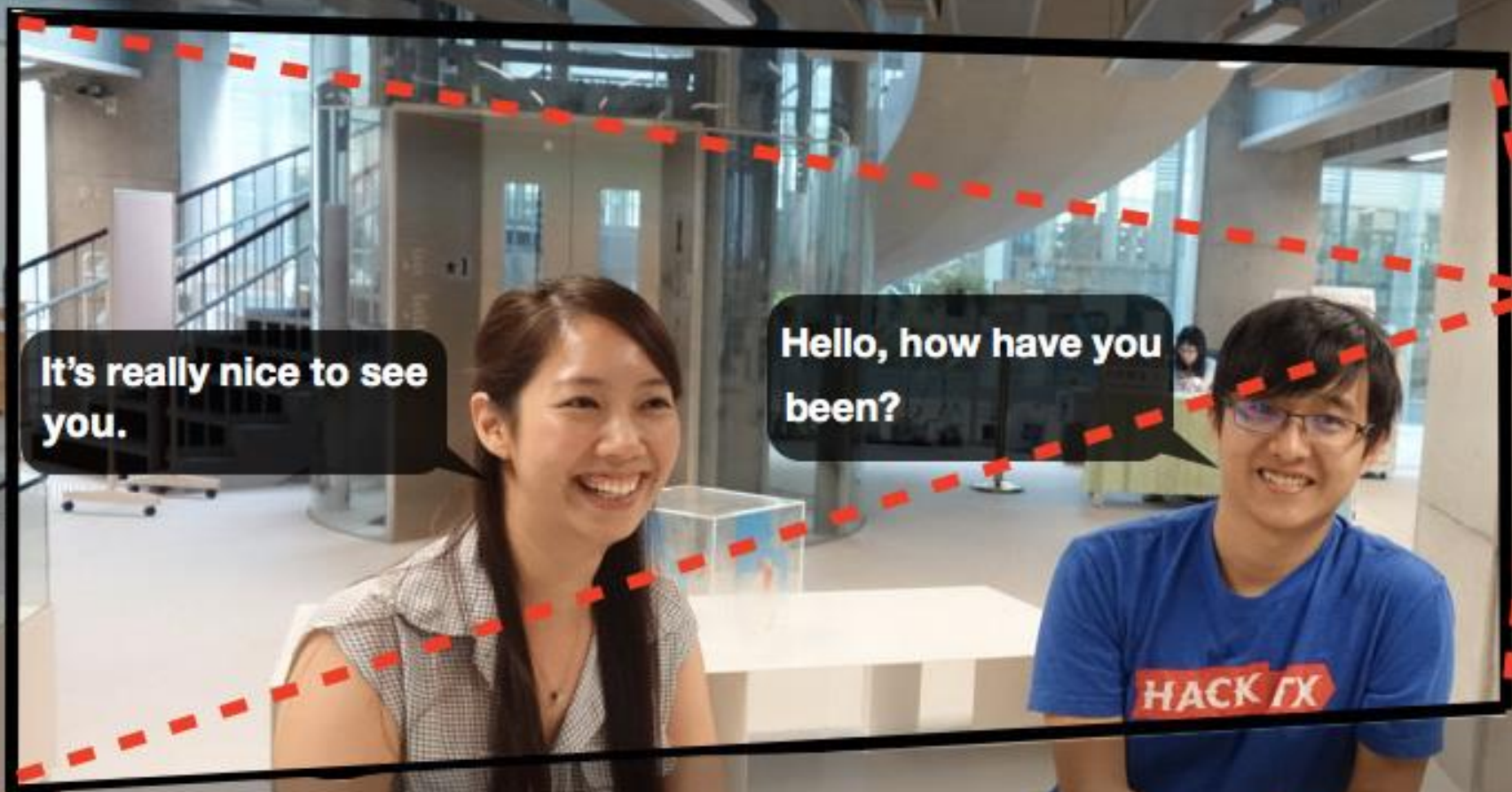


It's really nice to see you.



Hello, how have you been?

To reduce this visual split, researchers have used **HMD to show captions.**



It's really nice to see you.

Hello, how have you been?

HACK IX



However, no work has evaluated HMD-based captioning in a

mobile context.

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

Study 1



Study 1



Goal

- To assess the communication needs and potential technologies for DHH people in mobile contexts.

Study 1



Goal

- To assess the communication needs and potential technologies for DHH people in mobile contexts.

Participants

- 12 DHH individuals (5 males, 6 females, 1 did not disclose)
- Recruited through email, social media and snowball sampling

Study 1



Goal

- To assess the communication needs and potential technologies for DHH people in mobile contexts.

Participants

- 12 DHH individuals (5 males, 6 females, 1 did not disclose)
- Recruited through email, social media and snowball sampling

Study Method

- Two part **semi-structured formative interview in lab-setting**: (i) challenges in a mobile conversation, (ii) ideas for future captioning technology
- Three mobile scenarios were explicitly explored: walking, in transit, and recreational.

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

Background
and Past Work



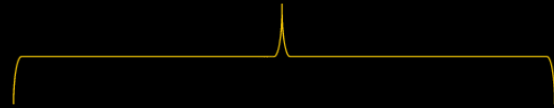
Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation



PART 1:

Mobile conversation
challenges

Background
and Past Work



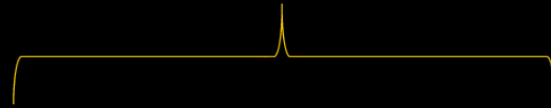
Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation



PART 1:

Mobile conversation
challenges

PART 2:

Captioning
technology design

Background
and Past Work



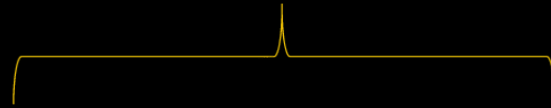
Study 1:
Interview



Proof-of-Concept
HMD Prototype




Study 2:
Evaluation



PART 1:

Mobile conversation
challenges



Participants had mobile conversations while **walking** to or from meetings, classes, and social activities as well as on **public transport** and in **cars**.

Participants had mobile conversations while **walking** to or from meetings, classes, and social activities as well as on **public transport** and in **cars**.

Assistive technology use in mobile contexts is **rare**.

A close-up photograph of a person's hands holding a smartphone. The image is dimly lit and has a dark, semi-transparent overlay. A quote is centered over the phone. The quote is in white text and is enclosed in large, stylized quotation marks. The background is a blurred, dark scene, possibly an indoor setting with other people present.

“I don't usually use technology other than hearing aids in moving conversations. I will occasionally use my phone to type something if it's impossible to hear. The [phone] isn't perfect because it demands that I split my attention and [also] have one [hand] holding the phone.”

MOVING CONVERSATION CHALLENGES

MOVING CONVERSATION CHALLENGES

- Conversations are **brief and shallow**

MOVING CONVERSATION CHALLENGES

- Conversations are **brief and shallow**
- **Hearing people** do not understand and accommodate needs



“If I need to look away for some reason, a deaf person will automatically stop talking and resume when I’m ready. A spoken conversation doesn’t have that type of natural stop and start...”

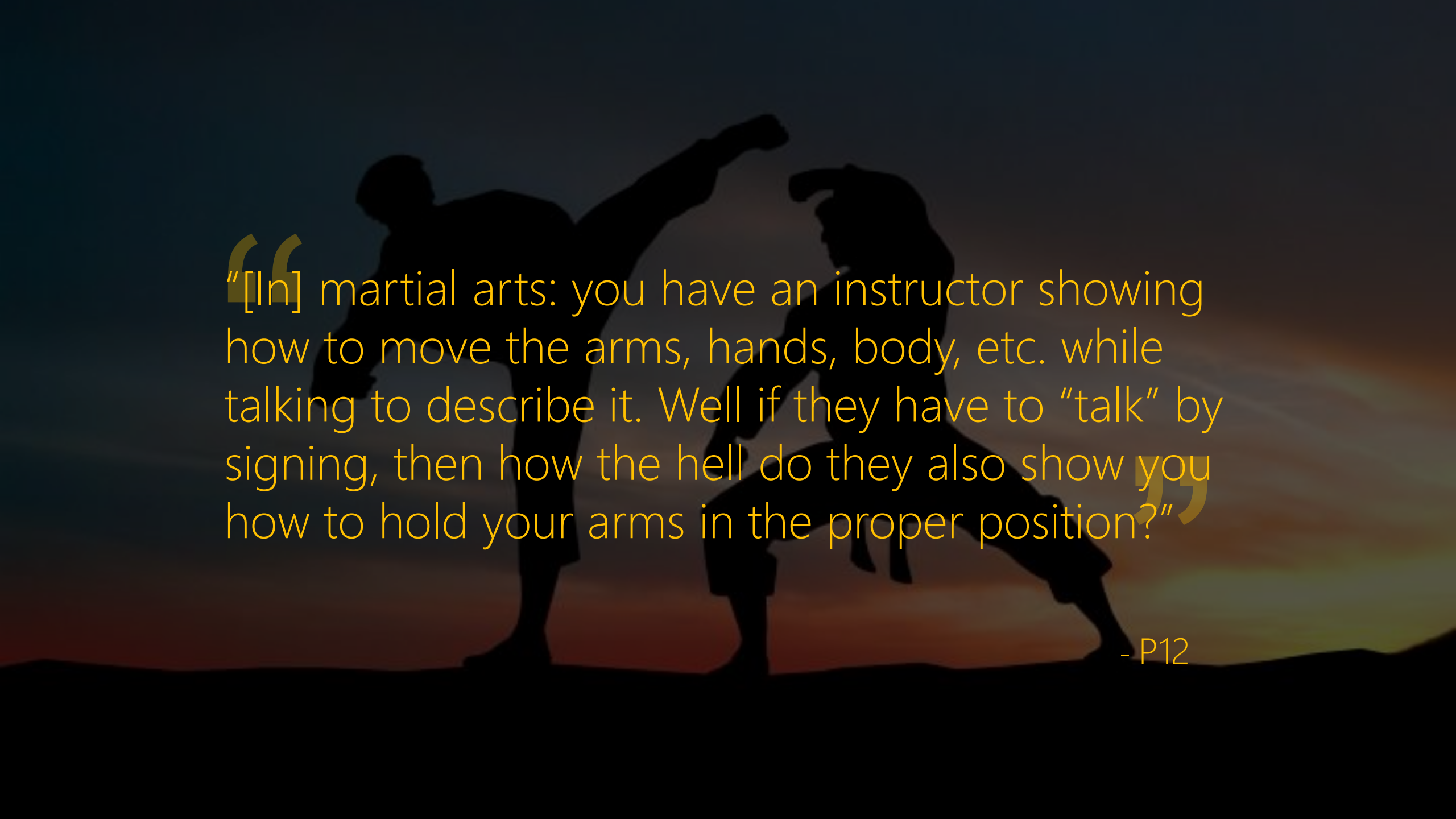
- P10

MOVING CONVERSATION CHALLENGES

- Conversations are **brief and shallow**
- **Hearing people** do not understand and accommodate needs

MOVING CONVERSATION CHALLENGES

- Conversations are **brief and shallow**
- **Hearing people** do not understand and accommodate needs
- **Recreational activities** are particularly challenging

The background of the slide features a sunset or sunrise scene with a gradient of colors from dark blue at the top to orange and red at the bottom. In the foreground, the silhouettes of two martial artists are visible. One figure on the left is in a low, ready stance, while the other on the right is performing a high kick, with their leg extended upwards and forward. The overall mood is dynamic and focused.

“[In] martial arts: you have an instructor showing how to move the arms, hands, body, etc. while talking to describe it. Well if they have to “talk” by signing, then how the hell do they also show you how to hold your arms in the proper position?”

- P12

MOVING CONVERSATION CHALLENGES

- Conversations are **brief and shallow**
- **Hearing people** do not understand and accommodate needs
- **Recreational activities** are particularly challenging

MOVING CONVERSATION CHALLENGES

- Conversations are **brief and shallow**
- **Hearing people** do not understand and accommodate needs
- **Recreational activities** are particularly challenging
- Challenges of varying **space, lighting** and **background noise**

Background
and Past Work



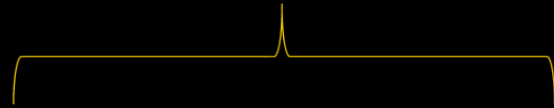
Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation



PART 1:
Characteristics
and Challenges

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

PART 1:
Characteristics
and Challenges

PART 2:
Captioning
Technology Design

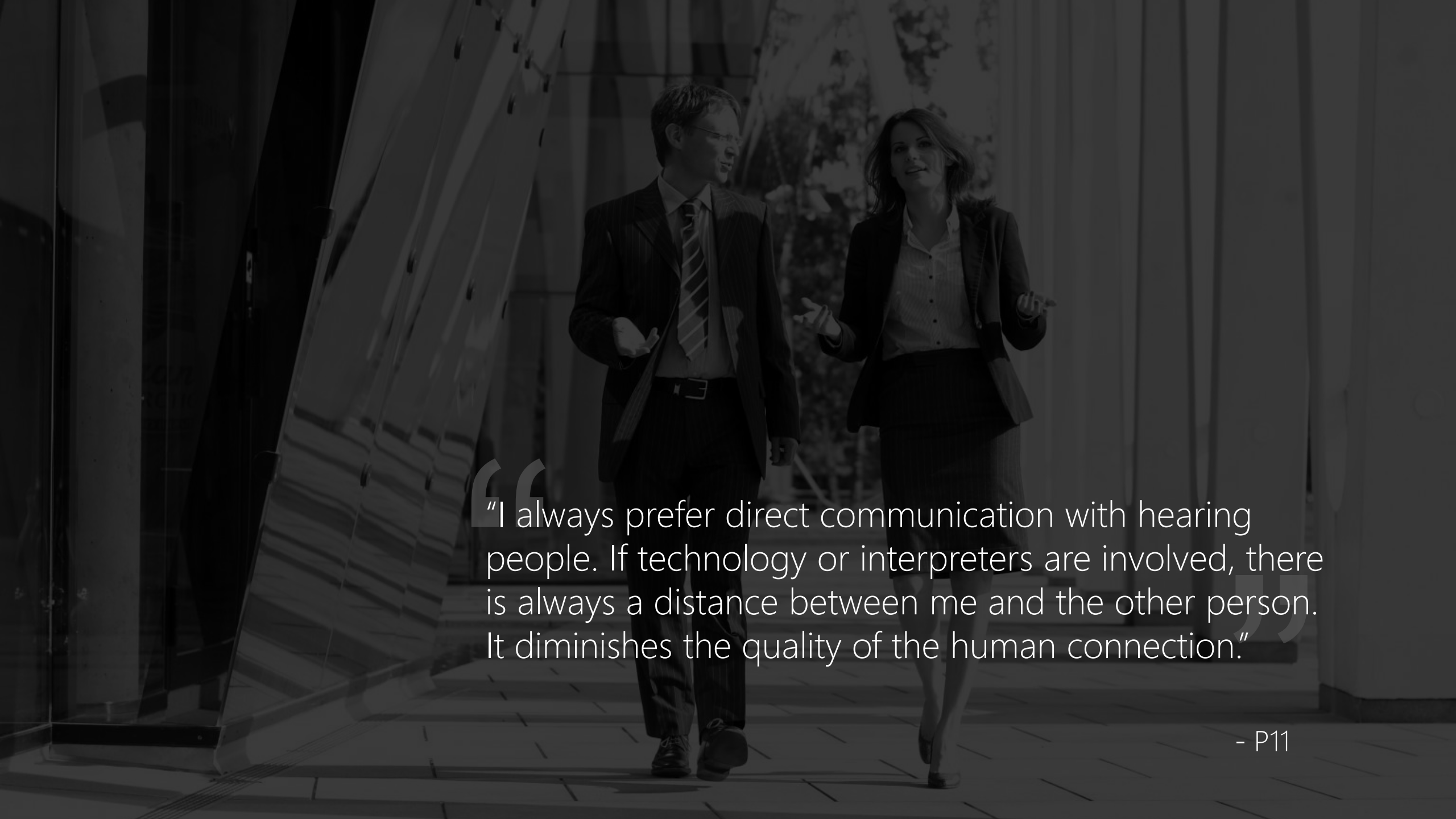
A family of three—a man, a woman, and a young boy—are walking away from the camera through a field of tall grass. The scene is set during sunset, with a warm, golden glow and long shadows. The man is on the left, the woman is in the middle, and the boy is on the right, holding the woman's hand. The background shows rolling hills under a clear sky.

All participants said they would use real-time captioning in at least one moving conversation scenario (walking, transit or recreational activity).

A family consisting of a man, a woman, and a young boy are walking away from the camera through a field of tall grass. The scene is set during sunset, with a warm, golden glow from the sun on the right side of the frame. The man is on the left, holding the hand of the woman on the right. The boy is in the foreground on the left, looking down. The background shows rolling hills under a clear sky.

All participants said they would use real-time captioning in at least one moving conversation scenario (walking, transit or recreational activity).

However, some were concerned that **captions may affect conversation quality** ($N=7$).

A man and a woman in business attire are walking and talking in a modern office hallway. The man is on the left, wearing a suit and glasses, and the woman is on the right, wearing a blazer and skirt. They are both smiling and gesturing with their hands. The background shows a large glass wall and a modern architectural design.

“I always prefer direct communication with hearing people. If technology or interpreters are involved, there is always a distance between me and the other person. It diminishes the quality of the human connection.”

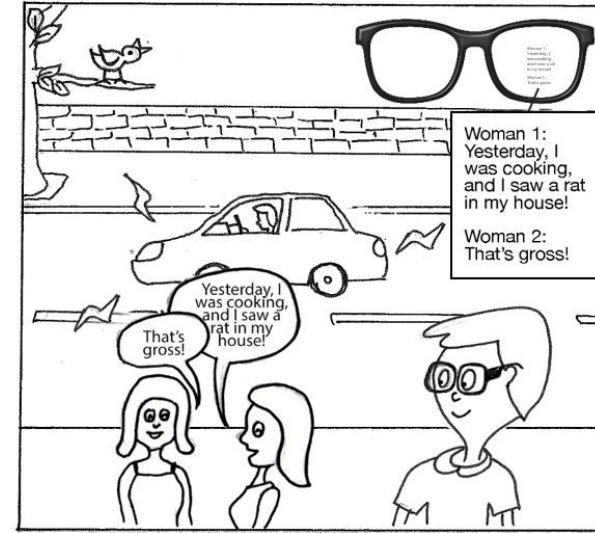
COMPARING DEVICES



SMARTPHONE



SMARTWATCH



HMD

	HMD	Smartphone	Smartwatch
Walking	$N=11$	$N=1$	
Transit (bus, car)	$N=6$	$N=4$	$N=1$
Recreational	$N=5$	$N=4$	

HMD was most preferred because it would **reduce the visual attention split.**

	HMD	Smartphone	Smartwatch
Walking	$N=11$	$N=1$	
Transit (bus, car)	$N=6$	$N=4$	$N=1$
Recreational	$N=5$	$N=4$	

	HMD	Smartphone	Smartwatch
Walking	$N=11$	$N=1$	
Transit (bus, car)	$N=6$	$N=4$	$N=1$
Recreational	$N=5$	$N=4$	

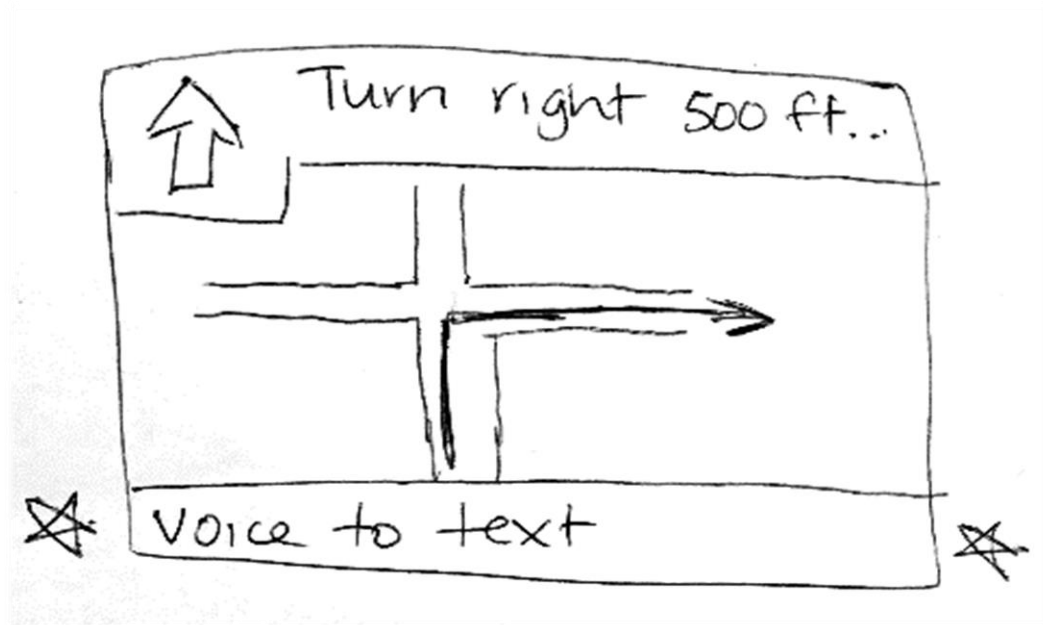
For high-contact sports, some people wanted smartphone because **HMD could fall off.**

Smartwatch was the least preferred because of the **small display size**.

	HMD	Smartphone	Smartwatch
Walking	$N=11$	$N=1$	
Transit (bus, car)	$N=6$	$N=4$	$N=1$
Recreational	$N=5$	$N=4$	

DESIGNS SKETCHED BY PARTICIPANTS

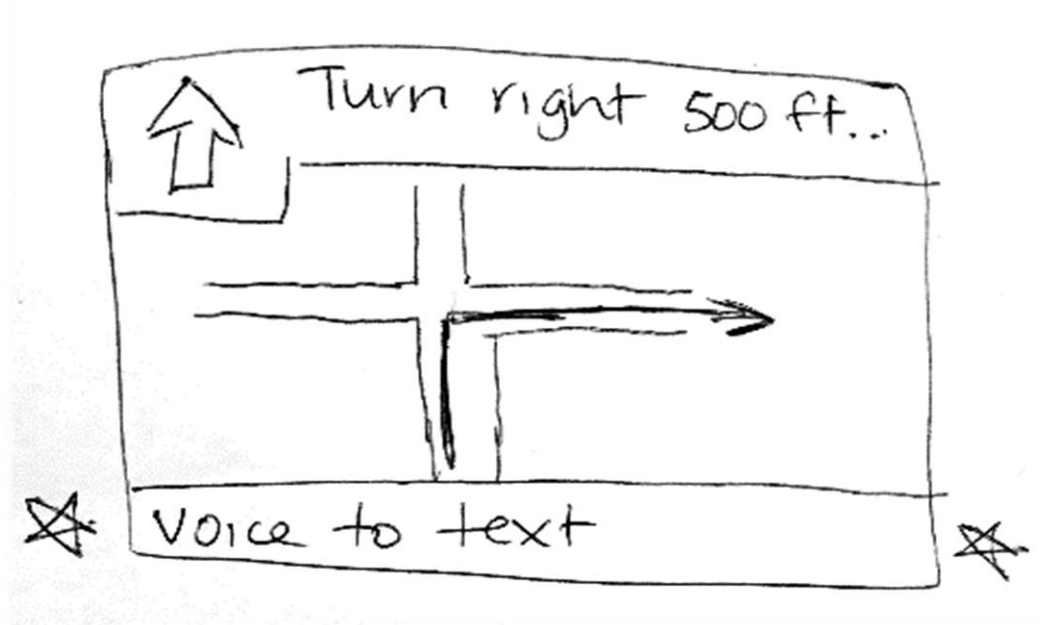
DESIGNS SKETCHED BY PARTICIPANTS



P5: Integrate captioning with car GPS

To reduce having to look at multiple devices

DESIGNS SKETCHED BY PARTICIPANTS



P5: Integrate captioning with car GPS
To reduce having to look at multiple devices



P2: A wrist worn device
To display captions

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

Background
and Past Work




Study 1:
Interview




Proof-of-Concept
HMD Prototype




Study 2:
Evaluation

A black and white photograph of a woman with long hair, wearing a VR headset and a patterned sweater with a dark scarf. She is smiling slightly. The background shows an office hallway with a window on the left and blinds on the right. A text box on the left contains a quote, with lines pointing to the headset and her face.


"This is third floor of the building.
We have a shop here.."

A black and white photograph of a woman with long hair, wearing a VR headset and a patterned sweater with a dark scarf. She is smiling slightly. The background shows an office hallway with a window on the left and blinds on the right. A text box is overlaid on the left side of the image, with lines pointing to the VR headset and the woman's face.

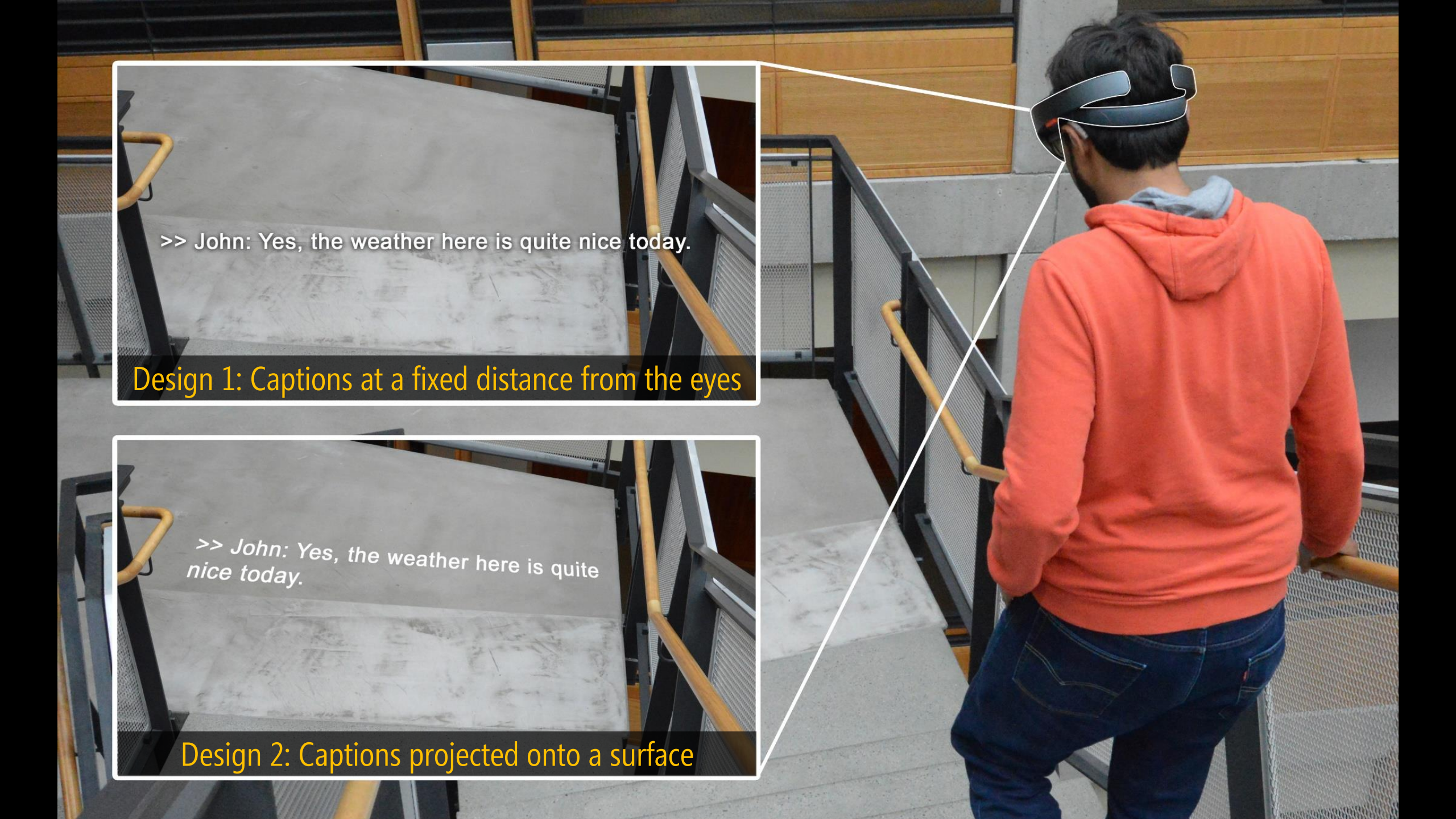
"This is third floor of the building.
We have a shop here.
Below, there's a restroom. "

A black and white photograph of a woman with long hair, wearing a VR headset and a dark scarf over a patterned sweater. She is smiling slightly. The background shows an office environment with a window and blinds. A speech bubble is overlaid on the left side of the image, pointing to the VR headset.

"This is third floor of
the building. We
have a shop here. "



"This is third floor of
the building. We
have a shop here. "

A person wearing an AR headset is shown from behind, standing on a staircase. Two white lines originate from the headset and point to two inset images. The top inset shows a caption at a fixed distance from the eyes, and the bottom inset shows a caption projected onto a surface.

>> John: Yes, the weather here is quite nice today.

Design 1: Captions at a fixed distance from the eyes

>> John: Yes, the weather here is quite nice today.

Design 2: Captions projected onto a surface

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

Background
and Past Work



Study 1:
Interview



Proof-of-Concept
HMD Prototype



Study 2:
Evaluation

Study 2



Study 2



Primary Goal

- To assess whether the use of HMD captions increased conversation accessibility, and decreased attention split for walking conversations

Study 2



Primary Goal

- To assess whether the use of HMD captions increased conversation accessibility, and decreased attention split for walking conversations

Participants

- 10 DHH individuals; 6 from Study 1

Study 2



Primary Goal

- To assess whether the use of HMD captions increased conversation accessibility, and decreased attention split for walking conversations

Participants

- 10 DHH individuals; 6 from Study 1

Part 1: Walking Scenario with HMD

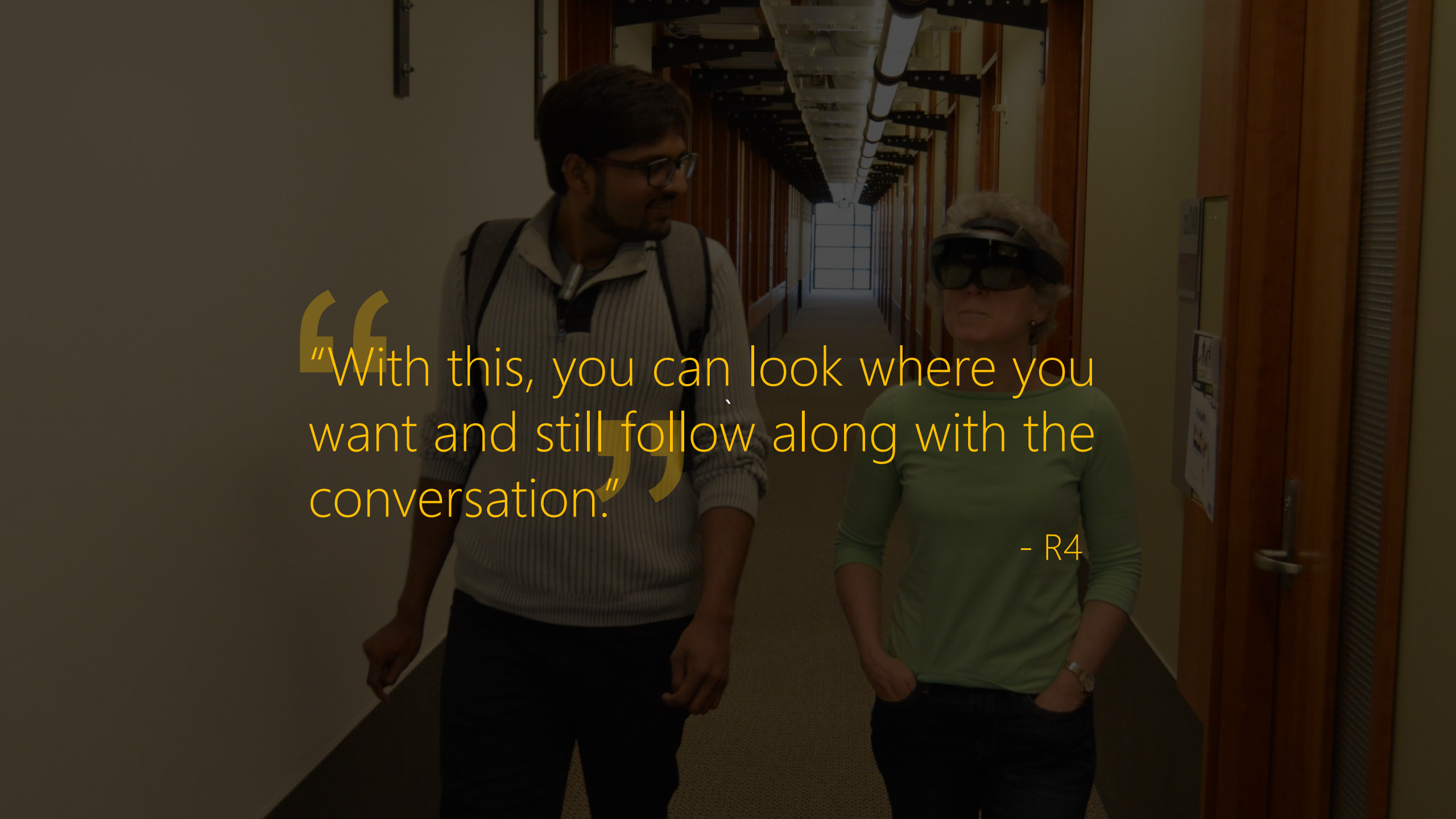
- One researcher conversed with the participant on casual topics
- Another researcher took observational notes
- The first researcher wore a lapel microphone that relayed speech to the on-site transcriber

Part 2: Open-ended Interview

- On the experience any feedback to the prototype

A photograph of a man and a woman walking down a long, brightly lit hallway. The man, on the left, is wearing a light-colored sweater and glasses. The woman, on the right, is wearing a green sweater, sunglasses, and a hat. They appear to be in conversation. The hallway has wooden walls and a carpeted floor. The image is dimmed to serve as a background for the text.


All participants used our prototype to understand at least some part of the conversation while walking.



“With this, you can look where you want and still follow along with the conversation.”

- R4

However, four participants found captions to be **occasionally distracting**.

A group of four men are standing outdoors. The man on the far left is wearing a light-colored hoodie and glasses. The man next to him is wearing a dark t-shirt with a logo that says "Research" and a VR headset. The man in the center is wearing a dark jacket. The man on the far right is wearing a dark t-shirt with the words "TEACHING ASSISTANT" printed on the back and a beanie. A quote is overlaid on the image, starting with a large opening quotation mark on the left and ending with a large closing quotation mark on the right.

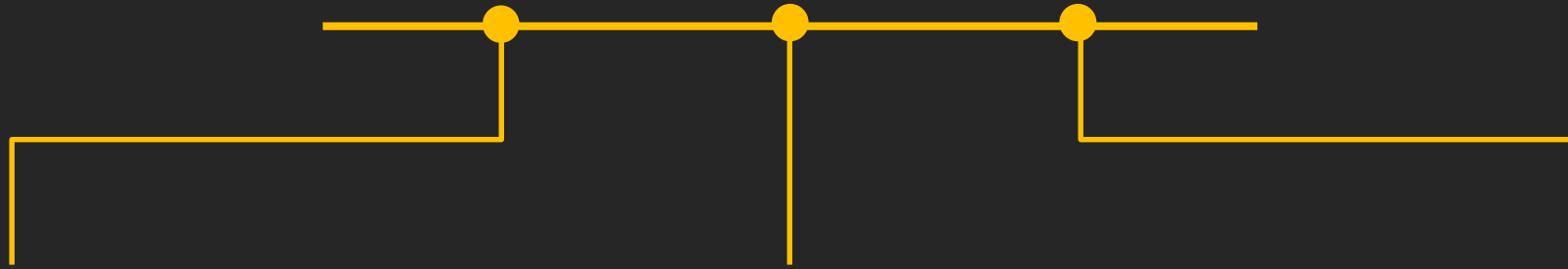
“When I was trying to formulate my own responses, I would find the captions quite distracting and, in cases like that, I wish [...] that I could look away from [the captions], at my discretion.”

- R5

High-Level Themes



High-Level Themes



High-Level Themes



Visual Split

- Participants used both speechreading and captions.
- Oral speakers looked at speakers more than captions. Sign language users focused on captions more.

High-Level Themes

```
graph TD; A[High-Level Themes] --- B[Visual Split]; A --- C[Caption Placement];
```

Visual Split

- Participants used both speechreading and captions.
- Oral speakers looked at speakers more than captions. Sign language users focused on captions more.

Caption Placement

- Split between whether to show captions in the field of view ($N=6$) or above speakers ($N=4$).
- All wanted the ability to turn off the captions when needed.

High-Level Themes

```
graph TD; Root[High-Level Themes] --- B1(( )); Root --- B2(( )); Root --- B3(( )); B1 --- V[Visual Split]; B2 --- C[Caption Placement]; B3 --- D[Design Suggestions];
```

Visual Split

- Participants used both speechreading and captions.
- Oral speakers looked at speakers more than captions. Sign language users focused on captions more.

Caption Placement

- Split between whether to show captions in the field of view ($N=6$) or above speakers ($N=4$).
- All wanted the ability to turn off the captions when needed.

Design Suggestions

- Display speaker identification cues (*e.g.*, name, location).
- Display environmental sounds (*e.g.*, door opening).
- Display voice tone and volume.



Reflection



As the **first work** to explore communication challenges and technology design for DHH people in mobile context, we have shown that:

1 Mobile context offer **new challenges** and a **new unexplored space** for innovation.

1 Mobile context offer **new challenges** and a **new unexplored space** for innovation.

2 Assistive technology in mobile contexts need to be **carefully designed**.

DESIGN GUIDELINES for HMD-Captioning



DESIGN GUIDELINES for HMD-Captioning

Automatic depth alignment



DESIGN GUIDELINES for HMD-Captioning

Automatic depth alignment

Adapt to changing context



DESIGN GUIDELINES for HMD-Captioning

Automatic depth alignment

Adapt to changing context

Convey **contextual information**



DESIGN GUIDELINES for HMD-Captioning

Automatic depth alignment

Adapt to changing context

Convey **contextual information**

Customizable



THE TEAM



Dhruv Jain
PhD Student, CSE, UW
djain@uw.edu



Rachel Franz
PhD Student, HCDE, UW
franzrac@uw.edu



Leah Findlater
Assist. Prof., HCDE, UW
leahkf@uw.edu



Jackson Cannon
UG Student, CSE, UW
jackscan@uw.edu



Raja Kushalnagar
Prof., Gallaudet University
raja.kushalnagar@gallaudet.edu



Jon Froehlich
Assoc. Prof., CSE, UW
jonf@uw.edu

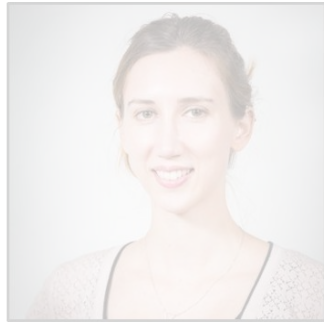
Towards Accessible Conversations in a Mobile Context

for People who are Deaf or Hard of Hearing

THE TEAM



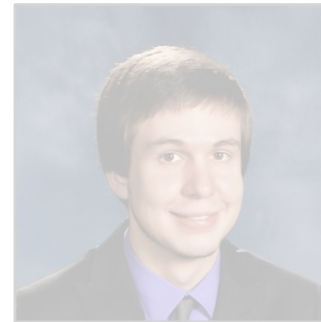
Dhruv Jain
PhD Student, CSE, UW
djain@uw.edu



Rachel Franz
PhD Student, HCDE, UW
franzrac@uw.edu



Leah Findlater
Assist. Prof., HCDE, UW
leahkf@uw.edu



Jackson Cannon
UG Student, CSE, UW
jackscan@uw.edu



Raja Kushalnagar
Prof., Gallaudet University
raja.kushalnagar@ga
llaudet.edu



Jon Froehlich
Assoc. Prof., CSE, UW
jonf@uw.edu

SPONSORS



UW
Reality Lab



Google
Faculty Research Awards



Microsoft