بسم الله الرحمن الرحیم

Information Technology Engineering

Mohammad Hossein Manshaei

manshaei@gmail.com

1401

Module A.2

Human Computer Interaction

Reference

- Mueller, F.F., Lopes, P., Strohmeier, P., Ju, W., Seim, C., Weigel, M., Nanayakkara, S., Obrist, M., Li, Z., Delfa, J. and Nishida, J., 2020, April. **Next Steps for Human-Computer Integration**. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-15).

Next Steps in Human-Computer Integration

Florian 'Floyd' Mueller [1*], Pedro Lopes [2*], Paul Strohmeier [3], Wendy Ju [4], Caitlyn Seim [5], Martin Weigel [6], Suranga Nanayakkara [7], Marianna Obrist [8], Zhuying Li [1], Joseph Delfa [1], Jun Nishida [2], Elizabeth M. Gerber [9], Dag Svanaes [10], Jonathan Grudin [11], Stefan Greuter [12], Kai Kunze [13], Thomas Erickson [14], Steven Greenspan [15], Masahiko Inami [16], Joe Marshall [17], Harald Reiterer [18], Katrin Wolf [19], Jochen Meyer [20], Thecla Schiphorst [21], Dakuo Wang [22], Pattie Maes [23]

Figure 1. Exemplars of Human-Computer Integration: extending the body with additional robotic arms; [70] embedding computation into the body using electric muscle stimulation to manipulate handwriting [48]; and, a tail extension controlled by body movements [86].

[1] Exertion Games Lab, Monash University, Melbourne, Australia.
[2] University of Chicago, Chicago, United States.
[3] University of Copenhagen, Copenhagen, Denmark and Saarland University, Saarbrücken, Germany.
[4] Cornell Tech, New York, Unites States.
[5] Stanford University, Stanford, United States
[6] Honda Research Institute Europe, Offenbach, Germany
[7] Augmented Human Lab, University of Auckland, Auckland, New Zealand.
[8] SCHI Lab, University of Sussex, Brighton, UK.
[9] Northwestern University, Evanston, Illinois, United States.
[10] Department of Computer Science, NTNU, Trondheim, Norway and IT University of Copenhagen, Denmark.
[11] Microsoft, Redmond, Washington, United States.
[12] Deakin University, Melbourne, Victoria, Australia.
[13] KMD, Keio University, Tokyo, Japan.
[14] Independent researcher, Minneapolis, Minnesota, United States.
[15] Strategic Research, CA Technologies, Pittsburgh, Pennsylvania, United States.
[16] University of Tokyo, Tokyo, Japan.
[17] Mixed Reality Lab, University of Nottingham, Nottingham, UK.
[18] University of Konstanz, Konstanz, Germany.
[19] Beuth University of Applied Sciences Berlin, Berlin, Germany.
[20] OFFIS-Institute for Information Technology, Oldenburg, Germany.
[21] School of Interactive Arts, Simon Fraser University, Vancouver, Canada.
[22] IBM Research, Cambridge, United States.
[23] MIT Media Lab, Cambridge, Massachusetts, United States.
* Authors contributed equally.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '20, April 25–30, 2020, Honolulu, HI, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-6708-0/20/04...15.00
DOI: 10.1145/3313831.3376242

ABSTRACT

Human-Computer Integration (HInt) is an emerging paradigm in which computational and human systems are closely interwoven. Integrating computers with the human body is not new. However, we believe that with rapid technological advancements, increasing real-world deployments, and growing ethical and societal implications, it is critical to identify an agenda for future research. We present a set of challenges for HInt research, formulated over the course of a five-day workshop consisting of 29 experts who have designed, deployed, and studied HInt systems. This agenda aims to guide researchers in a structured way towards a more coordinated and conscientious future of human-computer integration.

Author Keywords

Integration; augmentation; cyborg; implants; bodily extension; fusion; symbiosis

CCS Concepts

•**Human-centered computing → Interaction paradigms;**

INTRODUCTION

In designing the future of computing, it is no longer sufficient to think only in terms of the *interaction* between users and devices. We must also tackle the challenges and opportunities of *integration* between users and devices. This perspective is essential to fully understand and co-shape technology where user and technology together form a closely coupled system within a wider physical, digital, and social context.

Contents

- ➢ Human-Computer Integration
  - ✓ The Eras of Computing
  - ✓ Main Idea of HInt
- ➢ Low Vision
- ➢ EarBuddy

HCI vc HInt

- “How do we interact with computers?” towards “How are humans and computers integrated?”

The Eras of Computing

1. Mainframes (1945-)
2. Personal Computer(1980-)
3. Smart Phone(2007-)

Era / Paradigm	Users : Machines
Mainframe	many : 1
PC	1 : 1
Mobile	1 : many
Ubiquitous	many : many
Integration	**blurred boundary**

HInt

- Challenges of human-computer integration will help researchers and practitioners interested in HInt to:
  - (a) identify current knowledge, capabilities and areas of opportunity where they can contribute;
  - (b) situate their work within a larger HInt research agenda;
  - (c) and also allow policy makers to better understand the HInt community, state-of-the-art technology and research, as well as potential applications.

Contents

- ➢ Human- computer Integration
  - ✓ The Eras of Computing
  - ✓ Main Idea of HInt
- ➢ Low Vision
- ➢ EarBuddy

Idea of HInt

- The concept itself can be seen in science fiction, in concepts, such as
  - "man-machine mixture" in Edgar Allan Poe's writing in 1843
  - the humanoid-"robot" in Karel Capek's 1920s play;
  - in neuroscience where Manfred Clynes and Nathan Kline coined the term cyborg in the 1960s;
  - philosophy, as echoed in D. S. Halacy's 1965 essay on the Cyborg;
  - art, for example in Stelarc's 1990s work

- The integration is the future of human of computing. (Pattie Maes)
  - Inevitable
  - Necessary
  - desirable

Types of Human-Computer Integration

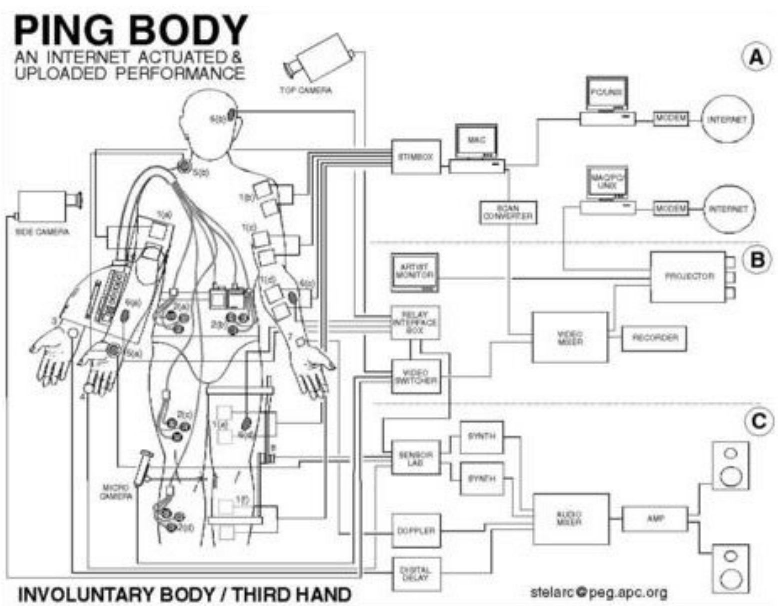
- These two dimensions map out a subset of ways in which humans and technology can relate, these are:
  - Symbiosis
  - Fusion

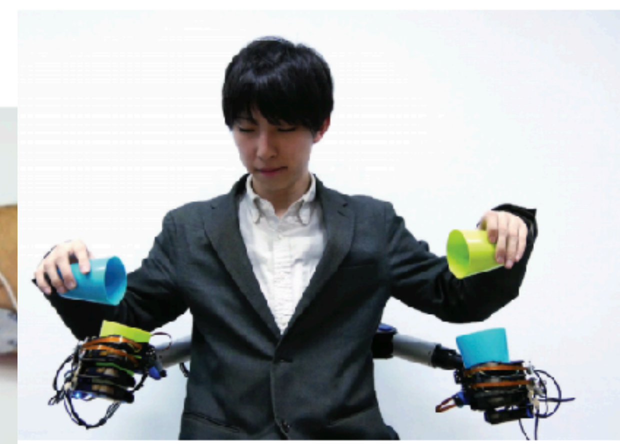
Symbiosis

- In this type of integration, agency is shared between humans and digital systems, and integration can occur on the individual level or between groups of people and technological systems.
- The key is that the agency is truly shared between technology and humans acting in concert, for example by collaborating in creative tasks or working together towards engaging experiences.

Fusion

- We define *fusion* as an integration in which devices extend the experienced human body or in which the human body extends devices





Is it useful after all!

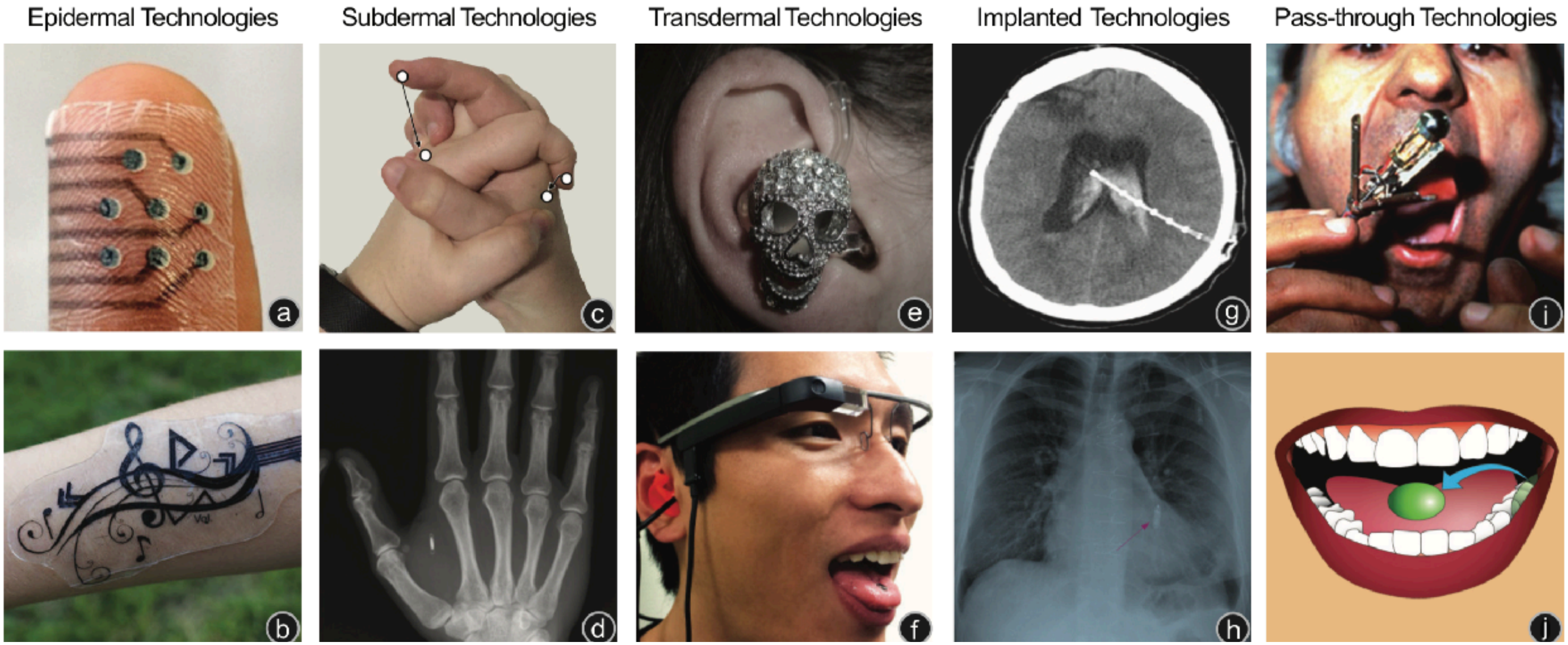
GRAND CHALLENGES OF HUMAN COMPUTER INTEGRATION

Grand challenges

- Four sets of challenges that we expect to be at the core of future HInt research:
  - (1) Human-Compatible Technology;
  - (2) Effects of Integration on Identity and Behavior;
  - (3) Integration and Society; and
  - (4) Designing Integrated Interaction.

Human-Compatible Technology

- Due to the fusion between computer and human body, HInt systems will benefit from a deeper understanding of the user's physiological and mental state.
- five key types of human-compatible technology
  - Epidermal technologies
  - Subdermal technologies
  - Transdermal technologies
  - Deep implanted technologies
  - Pass-through technologies



Type	Body Contact	Application	Permanence	Maintenance	Lifetime
Epidermal	Epidermis	User-controlled (sticker, spray-on)	Removable by the user	Removable and replaceable	User-controlled, allows for short term usage
Transdermal	Epidermis and dermis	Piercing or surgery	User-controlled or surgery	Through external port	Medium to long-term usage
Subdermal	Dermis	Syringe or small surgery	Permanence through surgery	Through surgery or wireless update	Medium to long-term usage
Deep Implanted	Internal organs	Surgery	Permanence through surgery	Through surgery or wireless update	Long-term usage
Pass-Through	Digestive system	User-controlled	No	Not intended: wait till it passes through	Usually 24-26 hours

Examples of human-compatible technologies

Materials for Integration

- Integration with the human body benefits from devices that feel and behave like parts of the body. For such devices, it is beneficial to be **biocompatible**, **miniaturized**, and **deformable**.

Effects of Integration on Identity and Behavior

- Perception of the Integrated Self: The relational self is the part of an individual's self-concept, which consists of the feelings and beliefs that one has regarding oneself and develops based on interactions with others
- Perception of other Integrated Selves
- Evaluating Potential Issues of the Self

Integration and Society

- Current and future HInt devices will affect society in a variety of ways.
- A list of key societal challenges:
  - Digital Divide
  - Body Bias
  - Mental and Physical Health
  - Ownership and Accountability

Designing Integrated Interaction

- HInt has two key qualities which are relevant to design:
  - the system can exhibit a form of autonomy that needs to be coordinated with the user and
  - the system's real-time feedback fuses with the user's sensations.
- Four key design challenges:
  - Integrating Novel Technologies
  - Designing Implicit Interaction
  - Designing for Variable Agency
  - Perceptual Transparency

Contents

- ➢ Human-Computer Integration
- ➢ Low Vision
  - ✓ Designing AR Visualizatio
  - ✓ VisualizationIs For Projection-Based AR
  - ✓ Visualizations for Smartglasses
- ➢ EarBuddy

Reference

- Zhao, Y., Kupferstein, E., Castro, B.V., Feiner, S. and Azenkot, S., 2019, October. **Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision**. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (pp. 387-402).

Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision

Yuhang Zhao[1], Elizabeth Kupferstein[1], Brenda Veronica Castro[1], Steven Feiner[2], Shiri Azenkot[1]

[1]Jacobs Technion-Cornell Institute, Cornell Tech, Cornell University, New York, NY, USA
{yz769, ek544, bvc5, shiri.azenkot}@cornell.edu

[2]Department of Computer Science, Columbia University, New York, NY, USA
feiner@cs.columbia.edu

ABSTRACT

Navigating stairs is a dangerous mobility challenge for people with low vision, who have a visual impairment that falls short of blindness. Prior research contributed systems for stair navigation that provide audio or tactile feedback, but people with low vision have usable vision and don't typically use nonvisual aids. We conducted the first exploration of augmented reality (AR) visualizations to facilitate stair navigation for people with low vision. We designed visualizations for a projection-based AR platform and smartglasses, considering the different characteristics of these platforms. For projection-based AR, we designed visual highlights that are projected directly on the stairs. In contrast, for smartglasses that have a limited vertical field of view, we designed visualizations that indicate the user's position on the stairs, without directly augmenting the stairs themselves. We evaluated our visualizations on each platform with 12 people with low vision, finding that the visualizations for projection-based AR increased participants' walking speed. Our designs on both platforms largely increased participants' self-reported psychological security.

Author Keywords

Accessibility; augmented reality; low vision; visualization.

ACM Classification Keywords

• **Human-centered computing~Mixed / augmented reality; Accessibility technologies.**

INTRODUCTION

As many as 1.2 billion people worldwide have *low vision*, a visual impairment that cannot be corrected with eyeglasses or contact lenses [11, 72]. Unlike people who are blind, people with low vision (PLV) have functional vision that they use extensively in daily activities [73, 74]. Low vision can be attributed to a variety of diseases (*e.g.*, glaucoma, diabetic retinopathy) and affects many visual functions including visual acuity, contrast sensitivity, and peripheral vision [21].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UIST '19, October 20–23, 2019, New Orleans, LA, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6816-2/19/10... $15.00
https://doi.org/10.1145/3332165.3347906

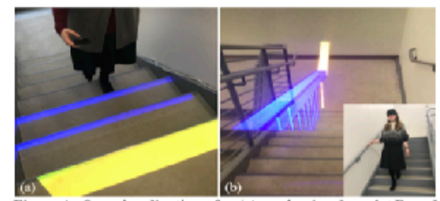


Figure 1: Our visualizations for (a) projection-based AR and (b) smartglasses to facilitate stair navigation for PLV.

Stair navigation is one of the most dangerous mobility challenges for PLV [5]. With reduced depth perception and peripheral vision [45, 56], PLV have difficulty detecting stairs or perceiving the exact location of stair edges [86]. As a result, PLV experience higher rates of falls and injuries than their typically-sighted counterparts [5, 13].

Despite the difficulty they experience, PLV use their residual vision extensively when navigating stairs [73]. Zhao *et al.* [86] found that they looked at contrast stripes (*i.e.*, contrasting marking stripes on stair treads) to perceive the exact location of stair edges; some also observed the trend of the railing to understand the overall structure of a staircase. However, sometimes stairs do not have contrast stripes, and even when they do, their stripes are often not accessibly designed; for example, stripes may have low contrast with the stairs or be too thin to detect [86]. Today, the only known tool to assist with stair navigation is the white cane, which many PLV prefer not to use [86]. Thus, there is a gap in tools that support PLV in the basic task of stair navigation.

Advances in augmented reality (AR) present a unique opportunity to address this problem. By automatically recognizing the environment with computer vision, AR technology has the potential to generate corresponding visual and auditory feedback to help people better perceive and navigate the environment more safely and quickly.

Our research explores AR visualization designs to facilitate stair navigation by leveraging PLV's residual vision. Designing visualizations for PLV is challenging [84, 85], especially for stair navigation, a dangerous mobility task. On one hand, the visualizations should be easily perceivable by PLV. A visualization that a sighted person can easily see (*e.g.*, a small arrow) may not be noticeable by PLV: it may be too small for them to see or outside their visual field [87]. On the other

Low Vision

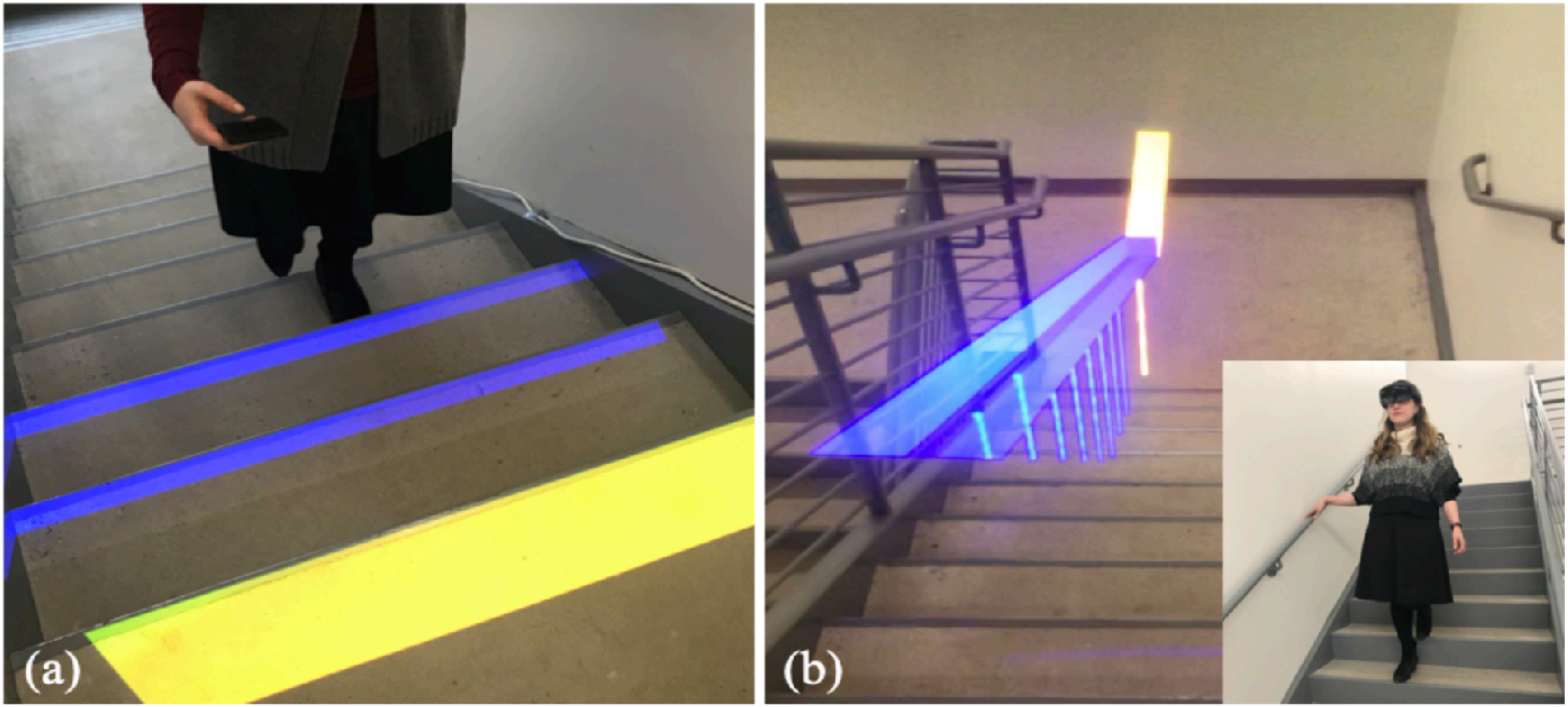
- As many as 1.2 billion people worldwide have low vision, a visual impairment that cannot be corrected with eyeglasses or contact lenses.
- Unlike people who are blind, people with low vision (PLV) have functional vision that they use extensively in daily activities.

Designing AR Visualizations

- Navigating stairs is a dangerous mobility challenge for people with low vision, who have a visual impairment that falls short of blindness.
- They conducted the first exploration of augmented reality (AR) visualizations to facilitate stair navigation for people with low vision. We designed
  - Visualizations for a projection-based AR platform
  - Smartglasses

Contents

- ➢ Human-Computer Integration
- ➢ Low Vision
  - ✓ Designing AR Visualization
  - ✓ VisualizationIs For Projection-Based AR
  - ✓ Visualizations for Smartglasses
- ➢ EarBuddy



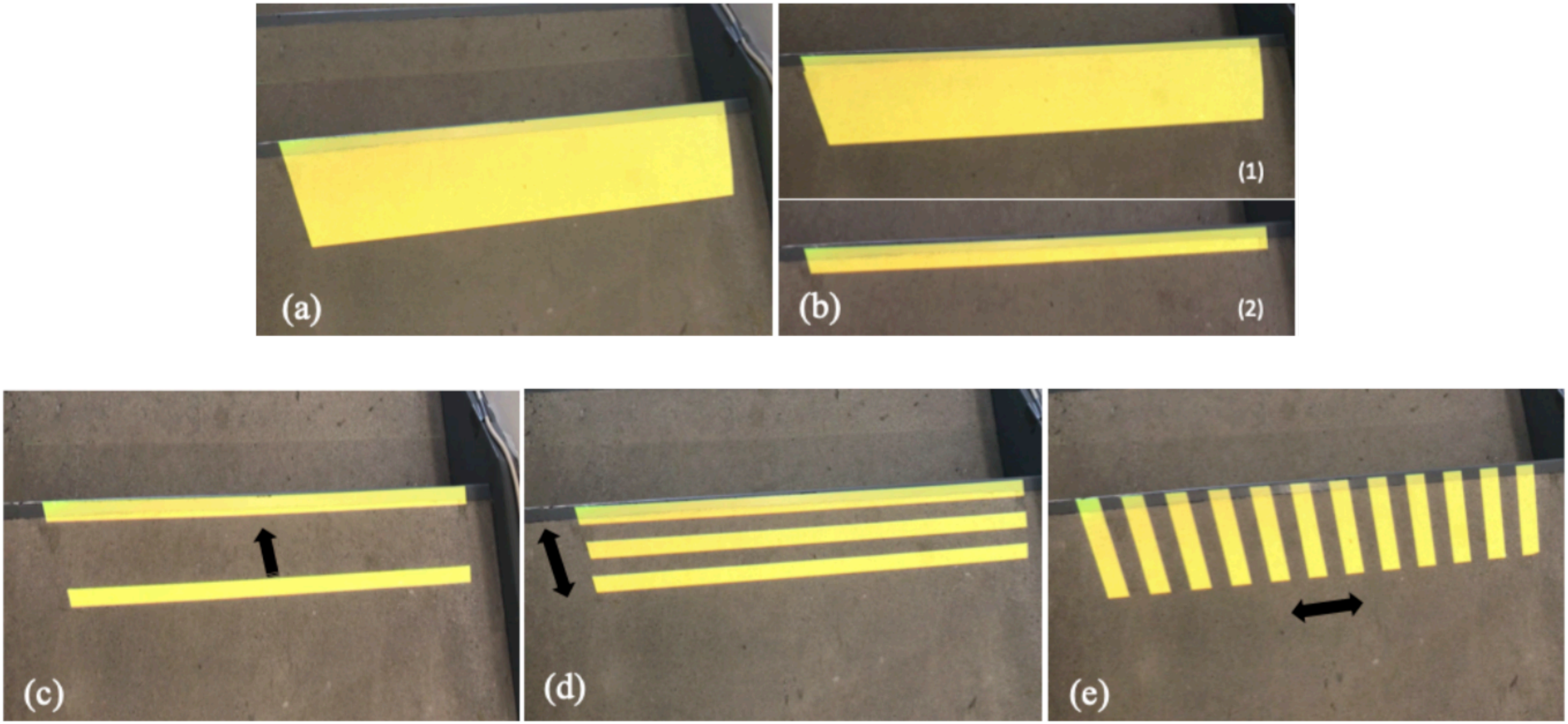
The visualizations for (a) projection-based AR and (b) smart glasses to facilitate stair navigation for PLV.

VisualizationIs For Projection-Based AR

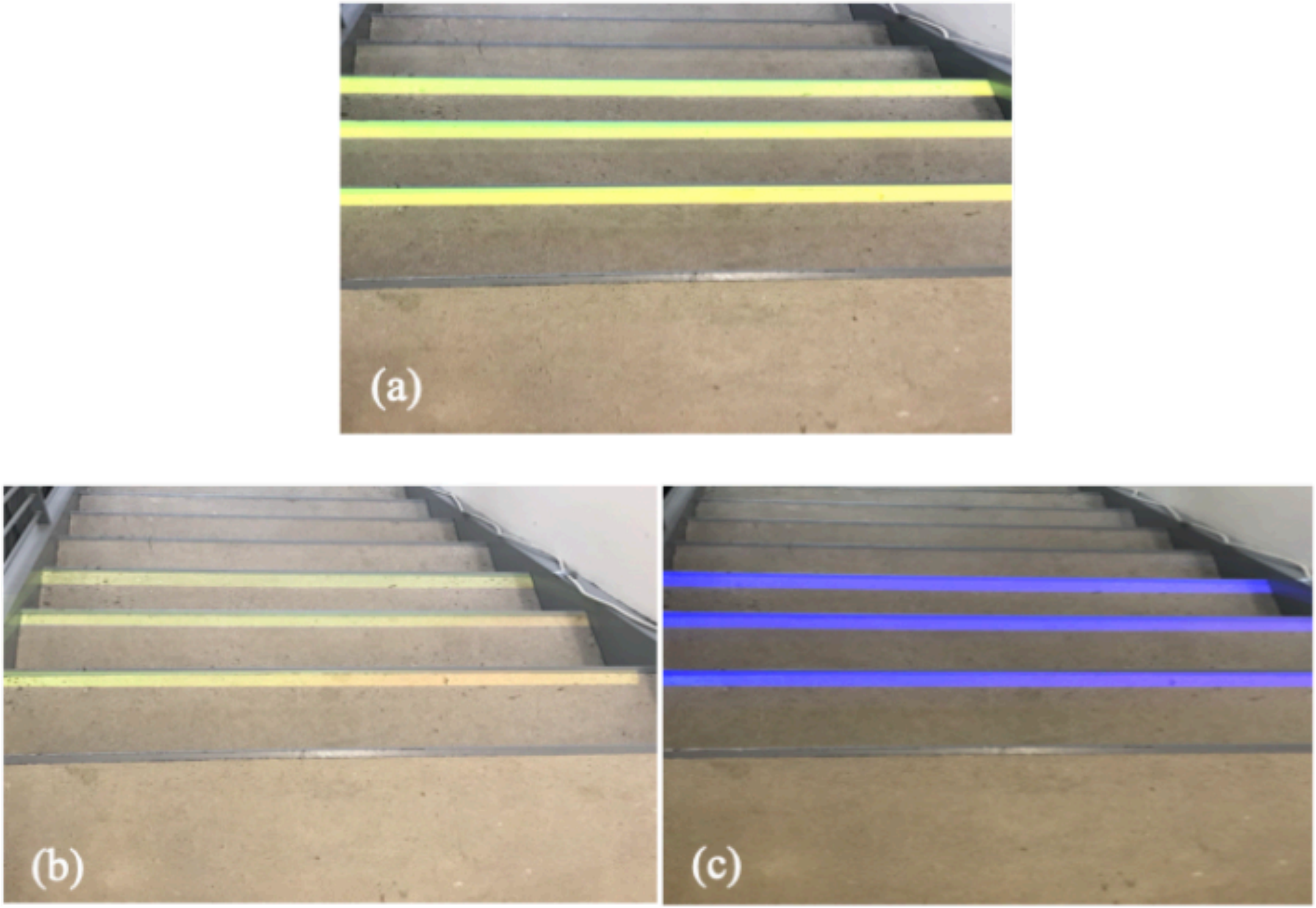
- We first explored the design **space of hand-held projection-based AR**, which combines a **camera** that recognizes the environment and a projector that projects visual contents into that environment.
- This platform has potential to **facilitate mobility** because it can project over **a relatively large area** and provide visual augmentations in people's peripheral vision, which is shown to be important for stair navigation.
- Although there are no popular commercial devices in the market, researchers have prototyped different hand-held projection-based AR platforms.s

Visualization Design

- Five animations was designed:
  - **Flash**: Since a flash can attract people's attention, they added this feature to the end highlights.
  - **Flashing Edge**: When the end highlight flashes, the user may lose track of the edge position when the highlight disappears.
  - **Moving Edge**: Movement also attracts attention. With a stable line at the stair edge, we added another line moving towards the edge to generate movement.
  - **Moving Horizontal Zebra**: Since movement can be distracting, we design a more subtle movement effect with a yellow and black zebra pattern moving back.
  - **Moving Vertical Zebra**: Moving the highlight over the edge of the stair may distort the perceived location of the edge, so we also designed a zebra pattern that is perpendicular to the edge.



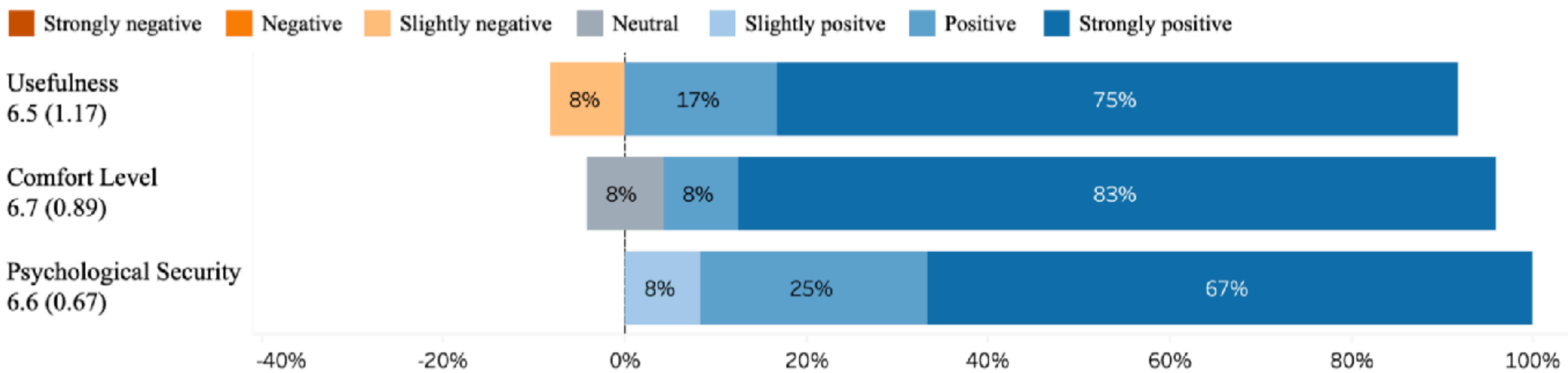
End highlights for first and last stairs. (a) Initial thick highlight with bright yellow; (b) Flashing Edge: the highlight switches between thick (b1) and thin (b2); (c) Moving Edge; (d) Moving Horizontal Zebra; (e) Moving Vertical Zebra.



Middle highlights: (a) Initial thin highlights with bright yellow; (b) Dull Yellow Highlights; (c) Blue Highlights.

Evaluation of Projection-Based AR Visualizations

- We evaluated the visualizations for projection-based AR, aiming to answer three questions:
  - <u>How do PLV perceive the different visualization designs?</u>
  - <u>How useful are the visualizations for stair navigation?</u>
  - <u>How secure do people feel when using our visualizations?</u>
- We recruited 12 PLV (6 female, 6 male; mean age=53.9) with different low-vision conditions, as shown (P1 – P12).



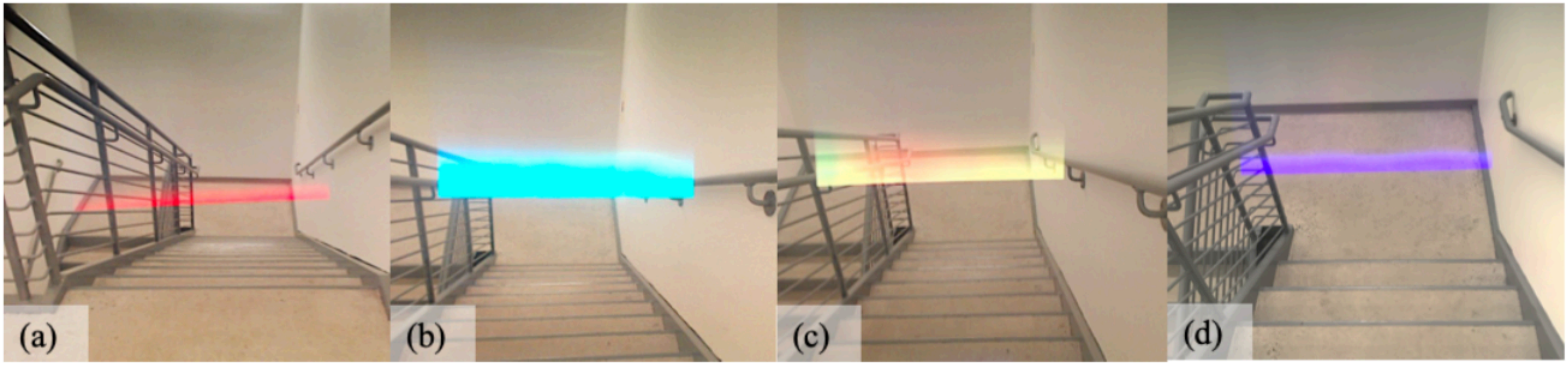
Diverging bars that demonstrate the distribution of participant scores for usefulness, comfort level, and psychological security when using visualizations on projection-based AR.

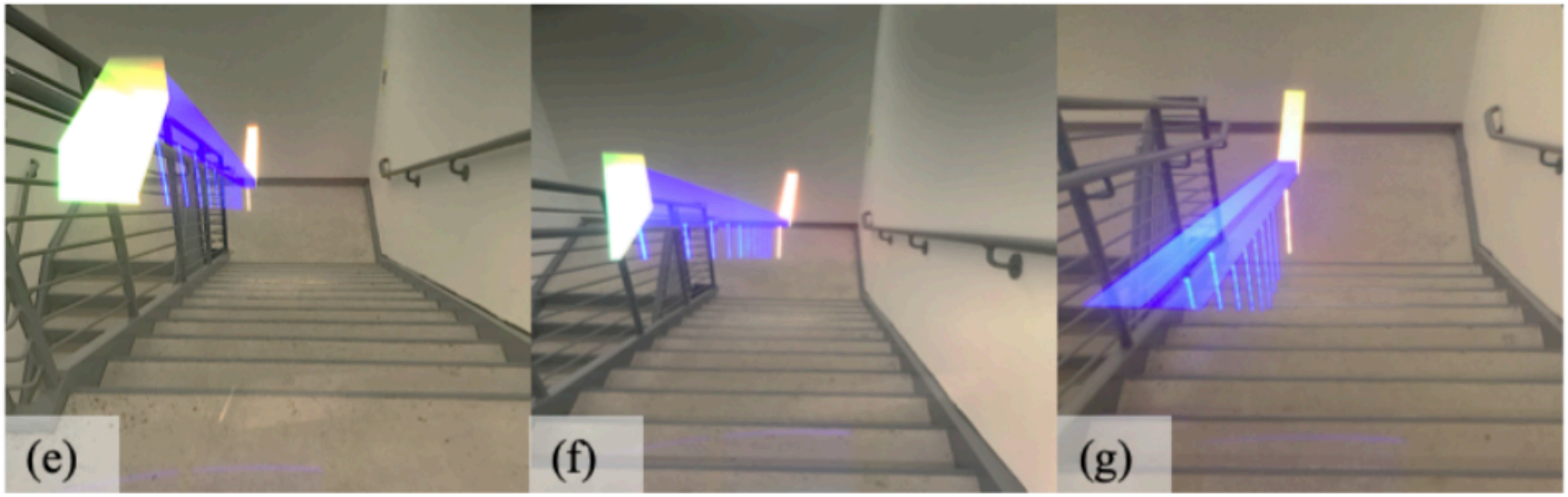
Contents

- ➢ Human-Computer Integration
- ➢ Low Vision
  - ✓ Designing AR Visualization
  - ✓ VisualizationIs For Projection-Based AR
  - ✓ Visualizations for Smartglasses
- ➢ EarBuddy

Visualizations for Smartglasses

- Current optical see-through smartglasses have a very limited field of view (FOV), largely limiting the area for presenting AR visualizations.
- While the recently announced HoloLens v2 is estimated to have a 29° vertical FOV, it is still much smaller than that of a typically-sighted human (120° vertical FOV).
- With the limited vertical FOV, the highlight design on projection-based AR would not work well for the smartglasses.
- To see the highlight on the current stair, a user would have to look nearly straight down to her feet, hindering her ability to see her surroundings.
  - This can be potentially dangerous and is physically strenuous.



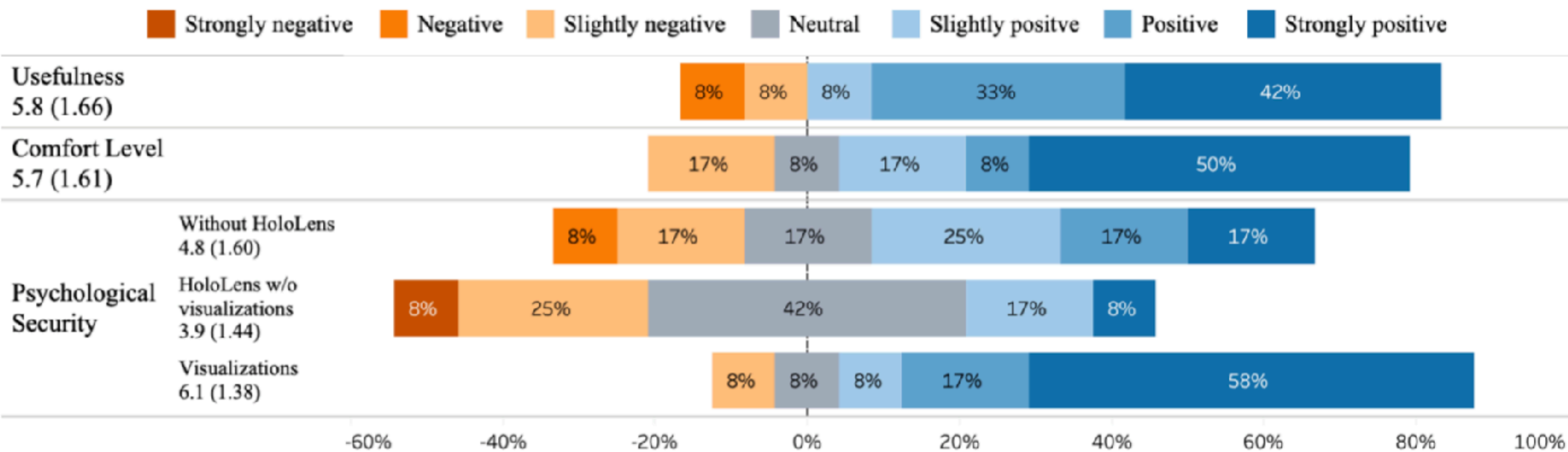


Glow (a–d) and Path (e–g). Glow: (a) thin red glow on the landing; (b) thick cyan glow in the preparation area; (c) thick yellow glow in the alert area; (d) thin blue glow on the middle of the stairs. Path: (e) view of the Path on the landing; (f) view of the Path when getting close to the first stair; (g) view of the Path on the middle of the stairs.

Evaluation of Smartglasses Visualizations

- How do PLV perceive the visualizations on smartglasses?
- How effective are the visualizations for stair navigation?
- How secure do PLV feel when using our visualizations?
- We recruited 12 PLV (5 female, 7 male; mean age=51.6) with different low vision conditions (P6– P17).

Distribution of participants' preferences for visualizations and sonification on HoloLens.



Diverging bars that demonstrate the distribution of participant scores (strongly negative 1 to strongly positive 7) for the usefulness and comfort level of the visualizations, and their psychological security in three conditions: without HoloLens, with Ho- loLens but no visualizations, and with visualizations.

Contents

- ➢ Human-Computer Integration
- ➢ Low Vision
- ➢ EarBuddy
  - ✓ Definition
  - ✓ EarBuddy Design
    - ✓ Gesture Selection
    - ✓ Data collection
    - ✓ Usability Evaluation
    - ✓ Potential Application

Reference

- Xu, X., Shi, H., Yi, X., Liu, W., Yan, Y., Shi, Y., Mariakakis, A., Mankoff, J. and Dey, A.K., 2020, April. **EarBuddy: Enabling On-Face Interaction via Wireless Earbuds**. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

EarBuddy: Enabling On-Face Interaction via Wireless Earbuds

Xuhai Xu[1,2], Haitian Shi[2,3], Xin Yi[2,4+], Wenjia Liu[5], Yukang Yan[2], Yuanchun Shi[2,4], Alex Mariakakis[3], Jennifer Mankoff[3], Anind K. Dey[1]

[1]Information School | DUB Group, University of Washington, Seattle, U.S.A

[2]Department of Computer Science and Technology, Tsinghua University, Beijing, China

[3]Paul G. Allen School of Computer Science & Engineering | DUB Group, University of Washington, Seattle, U.S.A

[4]Key Laboratory of Pervasive Computing, Ministry of Education, Beijing, China

[5]Department of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing, China

{xuhaixu, sht19, anind}@uw.edu, {yixin, shiyc}@mail.tsinghua.edu.cn, {sophie_liu}@bupt.edu.cn, {yyk15}@mails.tsinghua.edu.cn, {atm15, jmankoff}@cs.uw.edu

ABSTRACT

Past research regarding on-body interaction typically requires custom sensors, limiting their scalability and generalizability. We propose EarBuddy, a real-time system that leverages the microphone in commercial wireless earbuds to detect tapping and sliding gestures near the face and ears. We develop a design space to generate 27 valid gestures and conducted a user study (N=16) to select the eight gestures that were optimal for both human preference and microphone detectability. We collected a dataset on those eight gestures (N=20) and trained deep learning models for gesture detection and classification. Our optimized classifier achieved an accuracy of 95.3%. Finally, we conducted a user study (N=12) to evaluate EarBuddy's usability. Our results show that EarBuddy can facilitate novel interaction and that users feel very positively about the system. EarBuddy provides a new eyes-free, socially acceptable input method that is compatible with commercial wireless earbuds and has the potential for scalability and generalizability.

Author Keywords

Wireless earbuds; face and ear interaction; gesture recognition

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI);** *Interaction techniques; Ubiquitous and mobile computing systems and tools;*

INTRODUCTION

Past research from the human-computer interaction community has explored the use of surfaces on the body like the palms [65], arms [26], nails [27], and teeth [71] for convenient, subtle, and eyes-free communication [20]. Leveraging these surfaces has typically required custom sensors—fingertip cameras [60], ultrasonic wristbands [77], and capacitive fingernails [27], *etc.* Such custom sensors limit the scalability and generalizability to other applications.

+ indicates the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.

http://dx.doi.org/10.1145/3313831.3376836

Figure 1: EarBuddy leverages the microphone embedded in wireless earbuds to recognize gestures on the face or around the ears.

Our work takes advantage of the growing popularity of wireless earbuds as ubiquitous sensors for on-body sensing. Apple sold tens of millions of AirPods [16]. Other companies like Samsung [7] and Sony [8] are expected to show comparable trends in uptake of their earbuds. Although wireless earbuds are mainly used for audio output (*i.e.*, playing music and videos), most products also include a microphone for audio input so that people can respond to phone calls. The fact that wireless earbuds rest within a person's ears means that their microphone is conveniently situated near multiple surfaces that are suitable for on-body interaction: the cheek, the temple, and the ear itself. Tapping and sliding fingers across these surfaces generates audio signals that can be captured by an earbud, transmitted to a smartphone via Bluetooth, and then processed on-device to interpret gestures.

This observation gives rise to EarBuddy, a novel eyes-free input system that detects gestures performed along users' faces using wireless earbuds. As shown in Figure 1, users can easily control a music player or react to a notification by EarBuddy. Since EarBuddy augments the capabilities of devices that are already commercially available, our technique can easily be deployed through software updates to the phone to provide new interaction experiences for users.

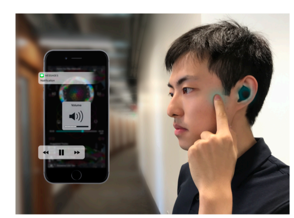
1

EarBuddy

- EarBuddy leverages the microphone embedded in wireless earbuds to **recognize gestures on the face or around the ears**.
- Wireless earbuds are mainly used for **audio output** (i.e., playing music and videos), most products also include a microphone for **audio input** so that people can respond to phone calls.
- EarBuddy is **a novel eyes-free input system** that detects gestures performed along users' faces using wireless earbuds.

EarBuddy

- Their contributions of this paper are threefold:
  - We propose EarBuddy, a novel eyes-free input technique supported by wireless earbuds **without the need for hardware modification**, and implement a real-time instantiation of EarBuddy.
  - We create a two-dimensional design space for **gestures near the face and ears**. Our first user study selects the gesture set for EarBuddy that is optimized for user preference and microphone detectability.
  - We train **a gesture recognition model** based on a second data collection study, and evaluate the usability of EarBuddy in a third user study.



EarBuddy

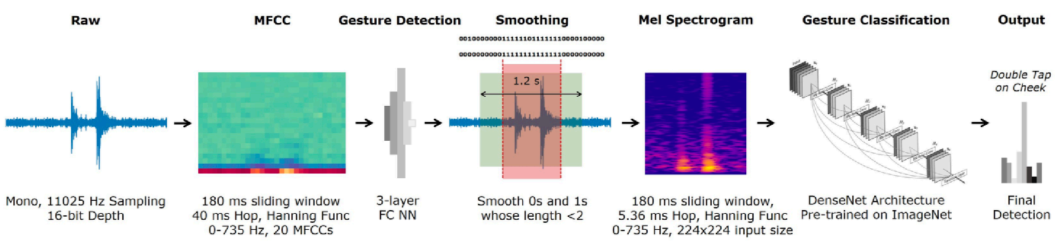
- They develop a comprehensive design space with 27 gestures along the side of a person's face and ears.
    - Since users cannot realistically remember all 27 gestures and some gestures are not easily detectable by earbud microphones, we conducted a user study (N=16) to narrow our gesture set to **eight gestures**. We carried out a second user study (N=20) to collect a thorough dataset with those gestures in both **a quiet environment and an environment with background noise**.
    - We used that data to train a shallow neural network binary classifier to detect gestures and **a deep DenseNet to classify gestures**. Our best classifier achieved a classification accuracy of 95.3%. Finally, we built a real-time implementation of EarBuddy using those models and conducted a third user study (N=12) to **evaluate EarBuddy's usability.**
- Their results show that EarBuddy sped up interactions by **33.9 - 56.2%** compared to touchscreen interactions. Users provided positive feedbac well, saying that EarBuddy can be used easily, conveniently, and naturally.

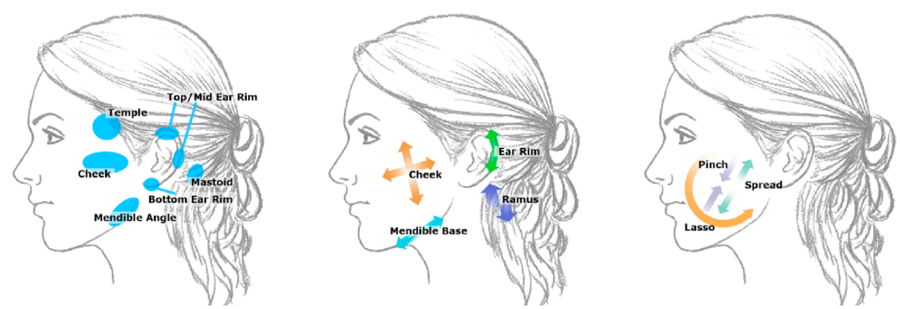
Contents

- ➢ Human-Computer Integration
- ➢ Low Vision
- ➢ EarBuddy
  - ✓ Definition
  - ✓ EarBuddy Design
    - ✓ Gesture Selection
    - ✓ Data collection
    - ✓ Usability Evaluation
    - ✓ Potential Application

EarBuddy Design

- We introduce both the sound-capturing system and interaction design below.
  - System Design
  - Interaction Design

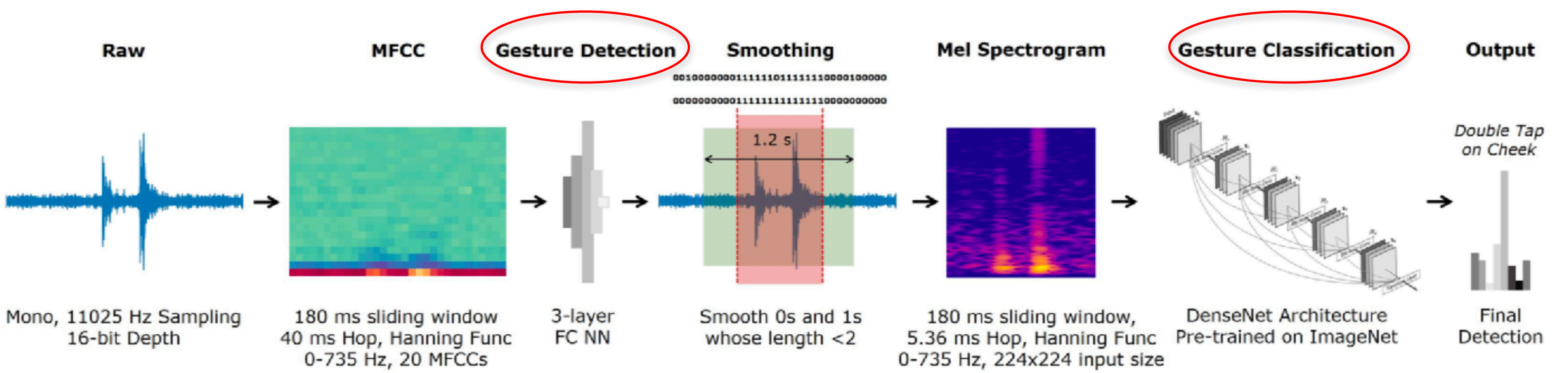




(a) Tap-based Gestures (b) Simple Slide-based Gestures (c) Complex Slide-based Gestures

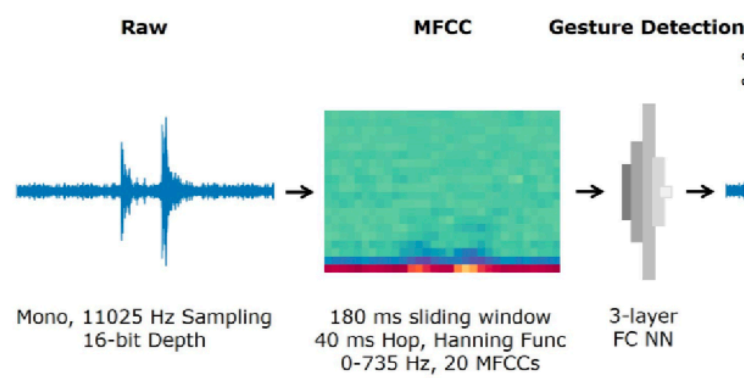
System Design

- EarBuddy recognizes gestures in two steps.
  - Detection
  - Classification

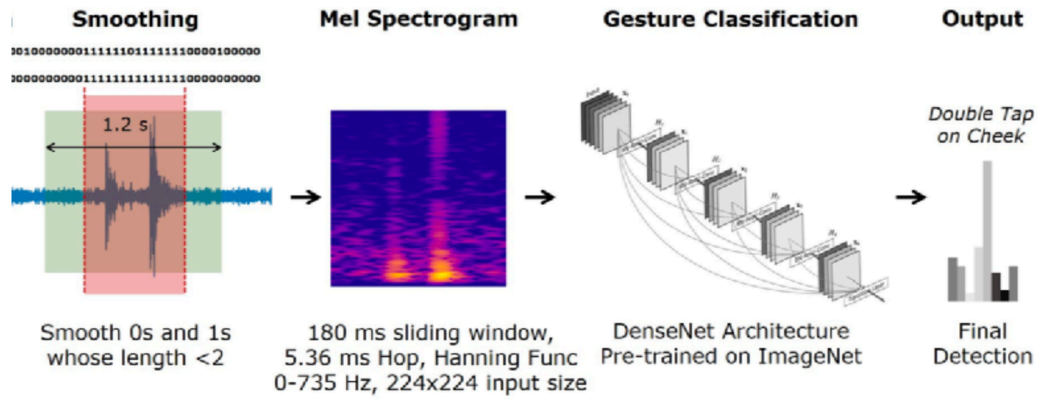


Detection

- Gesture detection starts using a 180 ms sliding window with a step size of 40 ms. Twenty MFCCs are extracted from the window at each step and fed into a binary neural network classifier.
- The classifier outputs
  - 1 — There is audio content belonging to a gesture
  - 0 — otherwise



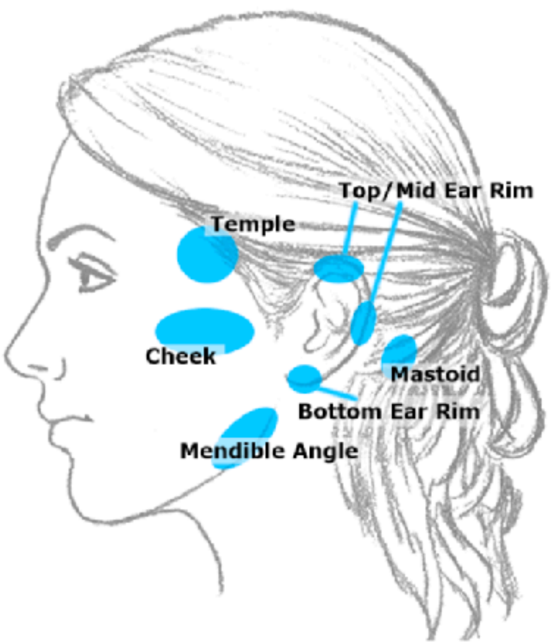
Classification



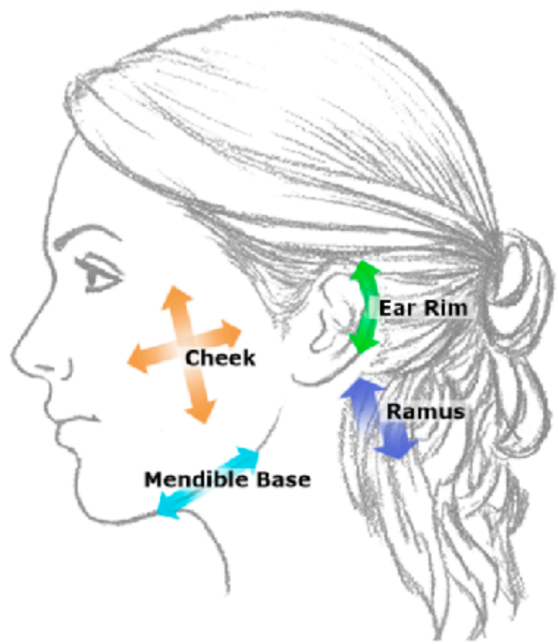
- Transfer learning using pre-trained vision models:
  - VGG16 - ResNet - DenseNet
- Modify this architecture after pre-training
  - Replacing the last fully-connected layer with **two fully-connected layers**,
  - using a **dropout layer a ReLU activation function** in between.
  - Modifying the output layer is required because EarBuddy requires far **fewer output classes**.
  - We train the modified, pre-trained network on our dataset to produce the final classification model used by EarBuddy.

Interaction Design

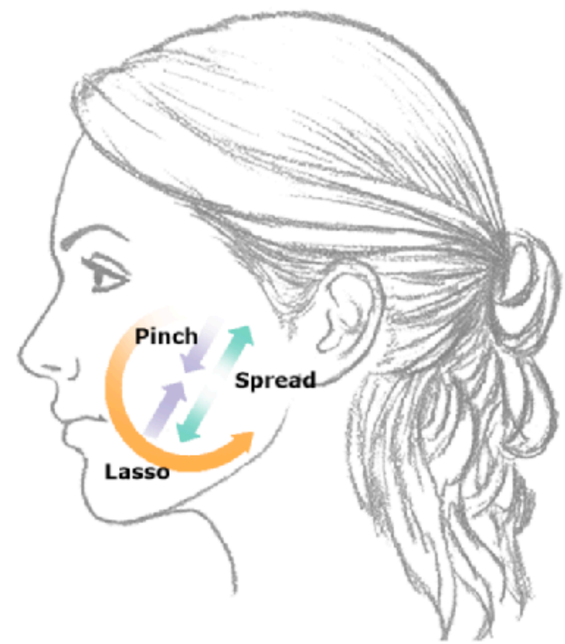
- People can produce different sounds by touching different areas around their face and ears. This is because the face and ears have unique structures with distinct combinations of materials.
- Using all possible pairs of options along those two dimensions that are feasible to perform, we generate 27 gestures.



(a) Tap-based Gestures



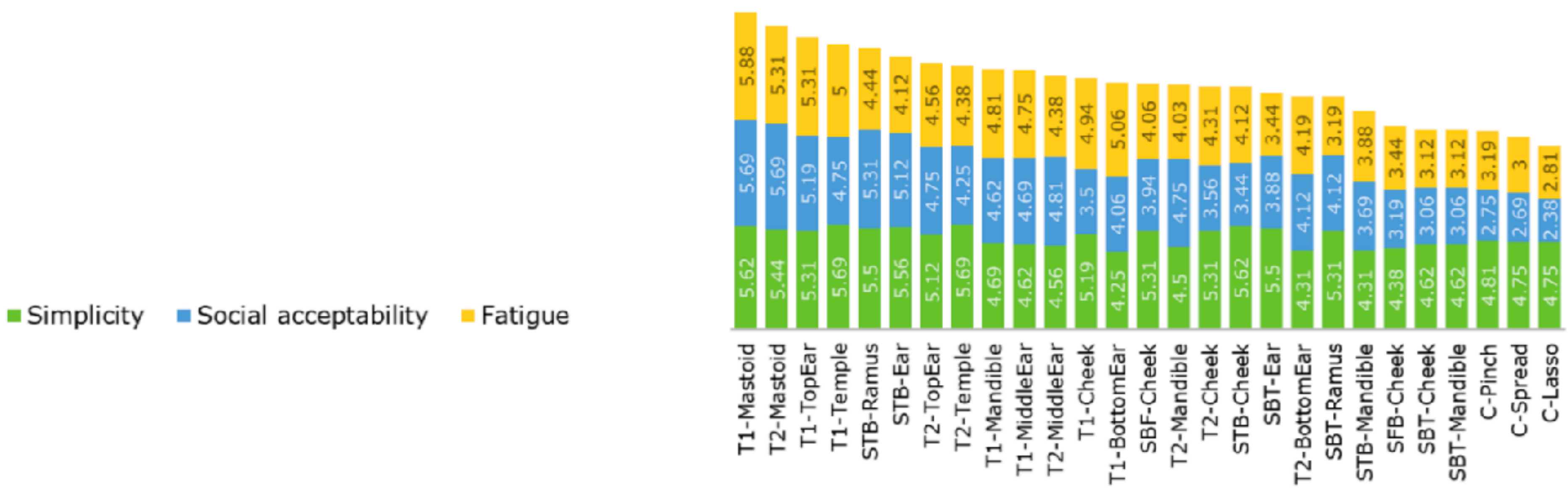
(b) Simple Slide-based Gestures



(c) Complex Slide-based Gestures

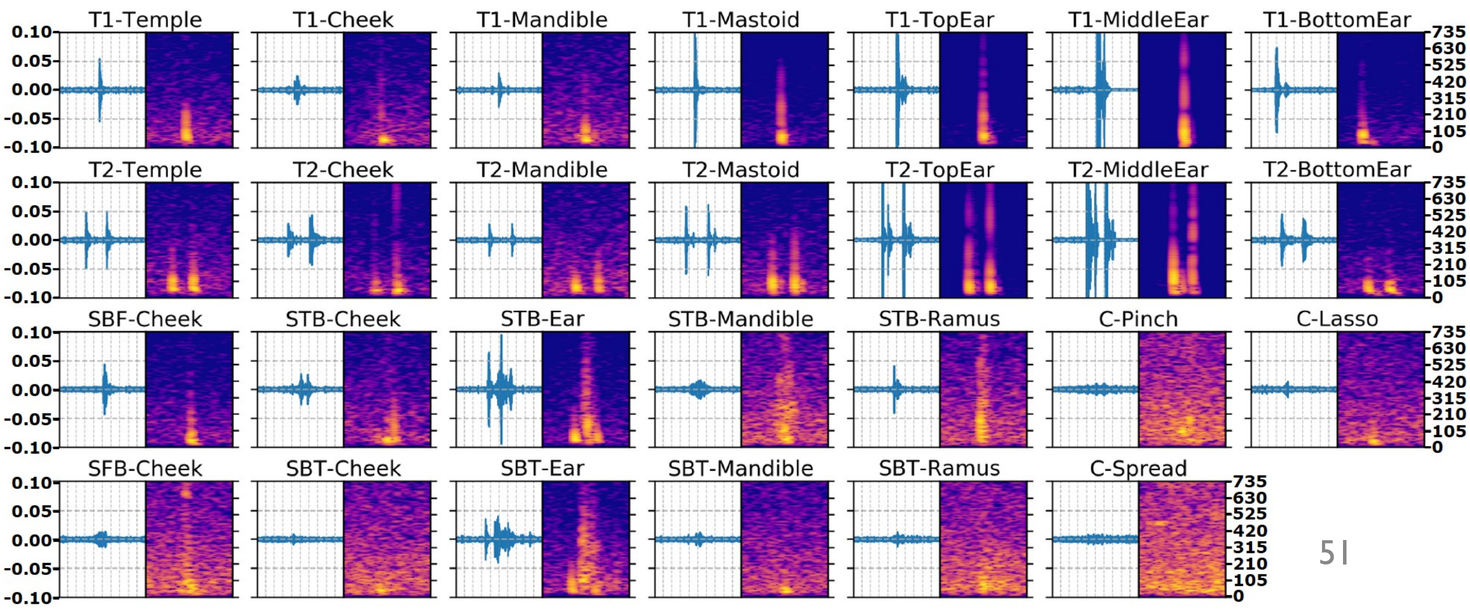
Gesture Selection

- 16 participants (8 male, 8 female, age = 21.3 ± 0.9)
- Each participant performed all 27 gestures three times using their right hand.
- After performing the gesture three times, the participant was asked to rate the gesture according to three criteria :
    - **Simplicity**: "The gesture is easy to perform precisely."
    - **Social acceptability**: "The gesture can be performed without social concern.
    - **Fatigue**: "The gesture makes me tired."



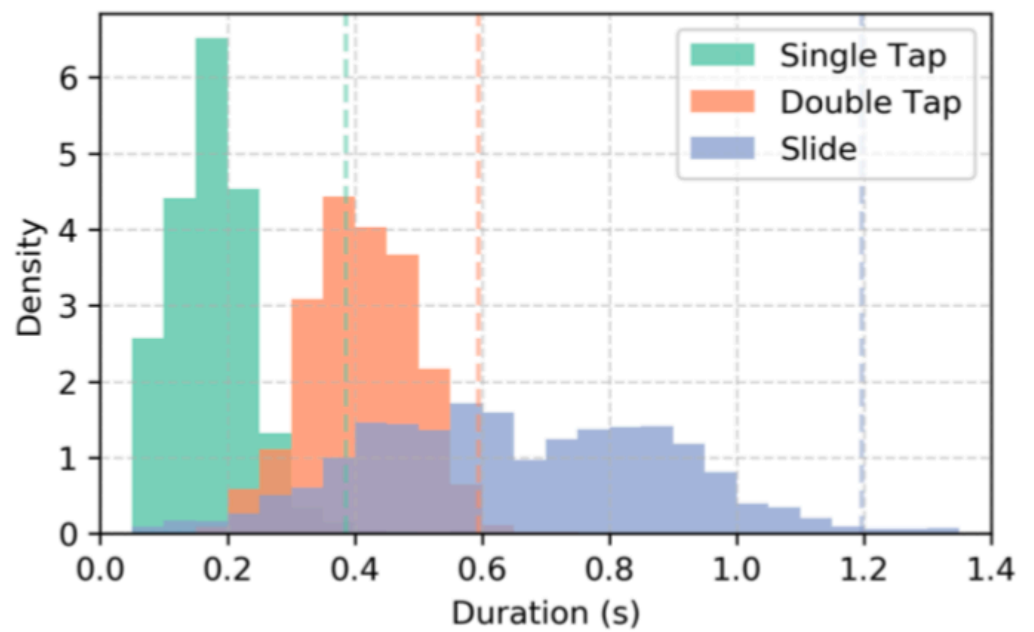
Gesture Selection

- We used the following aspects to select the **best gestures**:
  - Signal-to-noise ratio (SNR)
  - Signal Similarity
  - Design Consistency
  - Preference



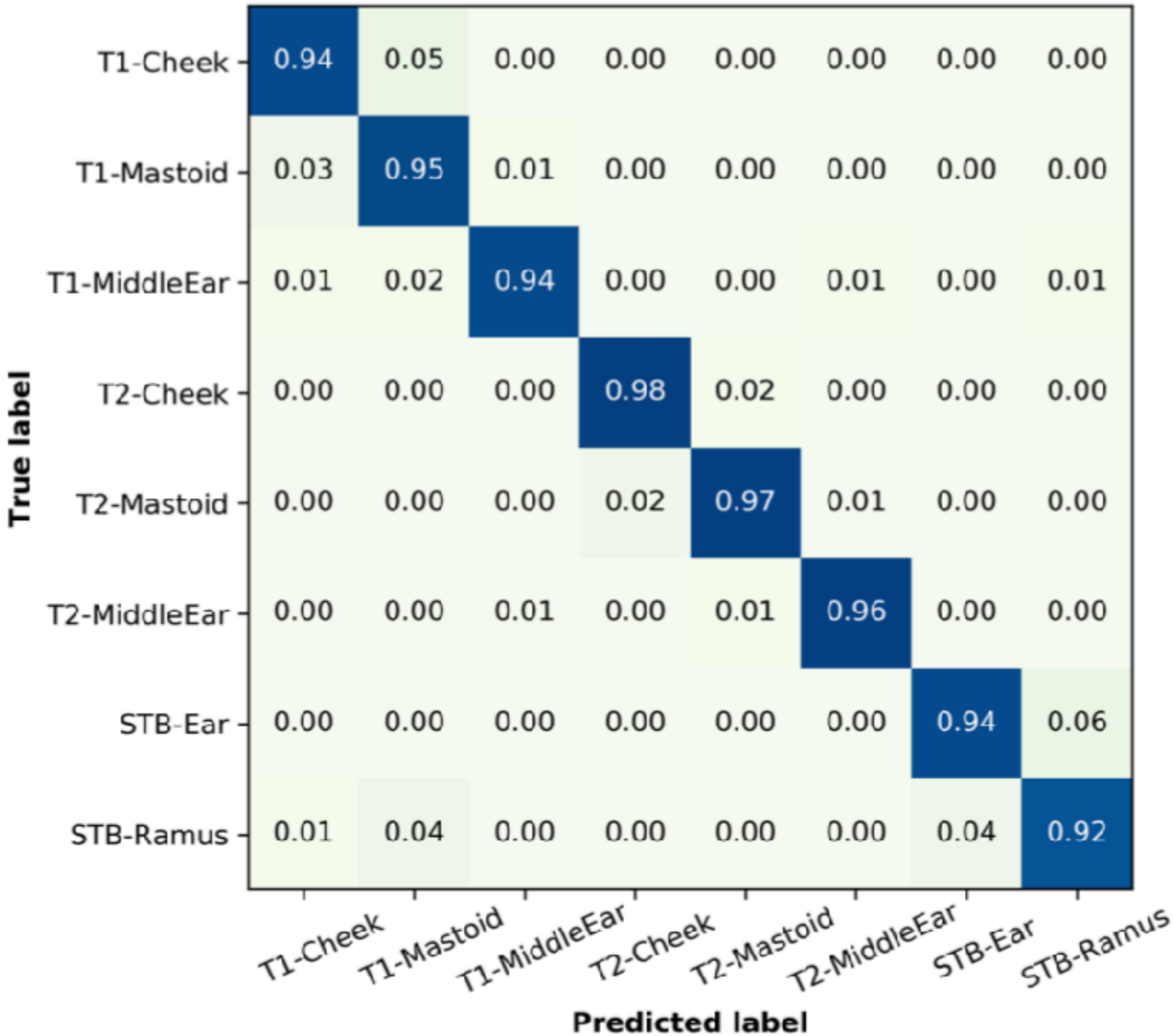
Data Collection

- 24 participants for this **studyNoisy Environment**
- Handling ambient noise with two sessions:
  - A quiet environment
  - A noisy environment
- The distribution of three gesture types' duration.



Model	Prec	Rec	F1	Acc
Random Forest on Means and Std of 20 MFCCs over the window	0.607	0.631	0.620	0.602
VGG16 from scratch with Adam	0.637	0.645	0.640	0.629
Pre-trained VGG16 with Adam	0.769	0.755	0.762	0.761
Pre-trained ResNet with Adam	0.810	0.793	0.802	0.785
Pre-trained DenseNet with Adam	0.807	0.803	0.805	0.809
Pre-trained DenseNet with Adam + Data Augmentation	0.872	0.872	0.872	0.872
Pre-trained DenseNet with SGD + Data Augmentation	0.929	0.893	0.916	0.914
Pre-trained DenseNet with SGD + Data Augmentation + Schedule	**0.956**	**0.951**	**0.954**	**0.953**

Test results of different models and enhancing techniques. Precision, recall, F1, and accuracy values are weighted across gestures.v



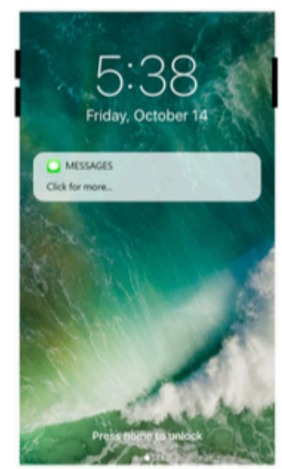
Confusion matrix of the best model on test set and Results with the leave-one-user-out data plus the ignored user's additional samples

Usability Evaluation

- Twelve participants (8 male, 4 female, age = 21.4 ± 0.8)
- Factors:
  - Setups: EarBuddy, Table, Pocket
  - Tasks: Music Player, Phone Call, Notifications

Music

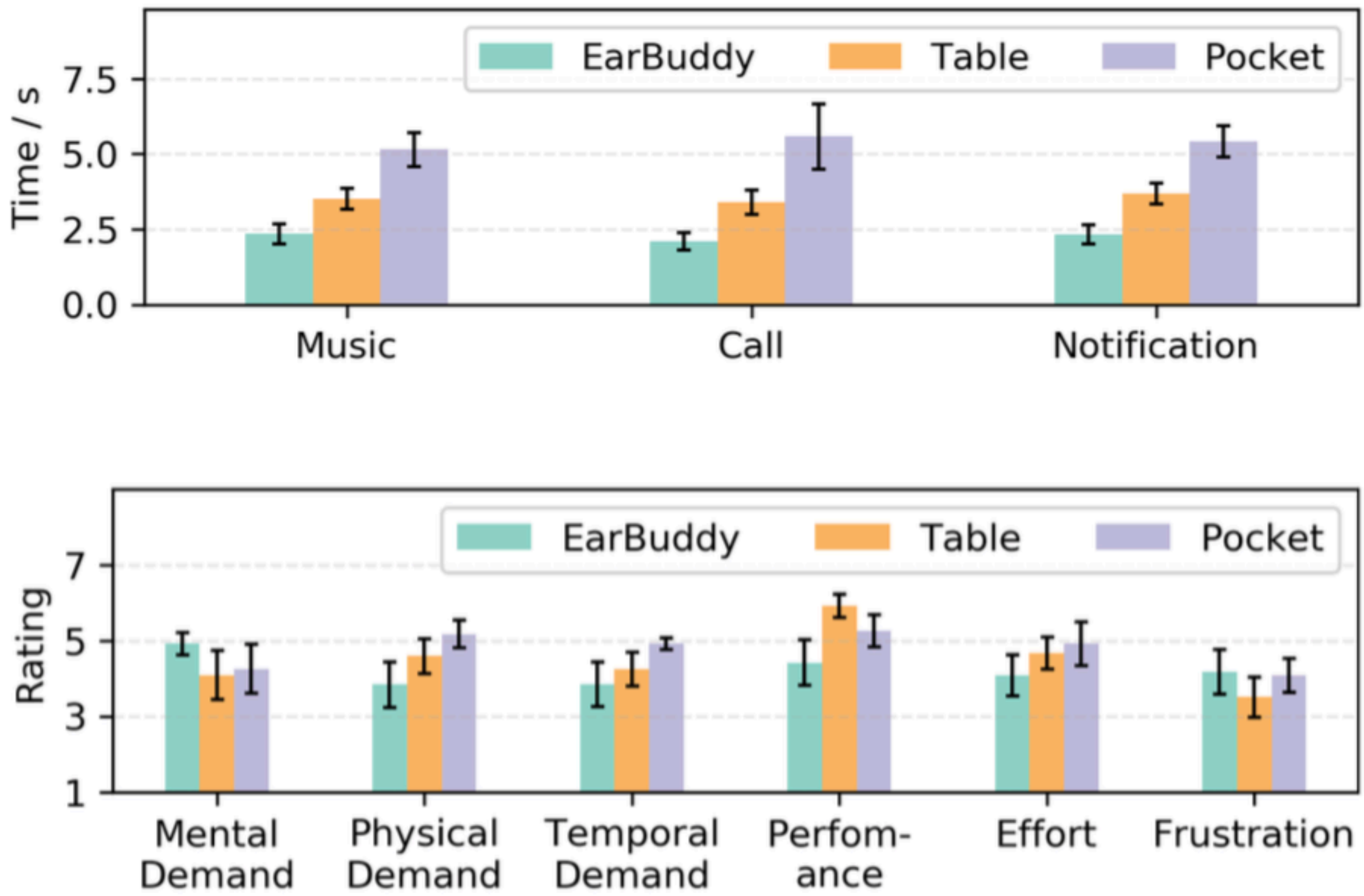
Call



Notification

Task	Operation	Earbud Gesture	Touch Gesture
Music	Play/Pause	STB-Ear	Virtual Button Click
Music	Vol Up	T1-Cheek	Physical Button Click
Music	Vol Down	T2-Cheek	Physical Button Click
Music	Next	T1-Mastoid	Virtual Button Click
Music	Previous	T2-Mastoid	Virtual Button Click
Call	Answer	T1-MiddleEar	Virtual Button Click
Call	Reject	T2-MiddleEar	Virtual Button Click
Call	Mute	STB-Ear	Physical Button Click
Notification	Read	STB-Ear	- (Read)
Notification	Open	T1-MiddleEar	Notification Click
Notification	Delete	T2-MiddleEar	Notification Slide

The design of the mapping of EarBuddy gestures and on-screen touch operations for the three applications examined in the user study.



Results of the evaluation study. Top) Time to complete the tasks. Bottom) Subjective ratings of the three setups

Potential Applications

- EarBuddy can provide an eyes-free, **socially acceptable** input method.
- Users can interact with devices in **a more subtle way**, e.g., during a meeting, in a library, and in an office.
- It is suitable for **quick reactions** such as issuing commands and handling notifications.
- EarBuddy can serve as **a convenient input method** when a user is using the device in a hands-free mode, such as when watching videos, cooking, etc.
- EarBuddy is **not suitable for repeated, continuous interactions**, e.g., text entry and interface scrolling.
- EarBuddy can be embedded in a headset **without additional hardware modification**.