Exploring Audio Icons for Content-Based Navigation in Voice User Interfaces

Jonas Kjeldmand Jensen & Daniel Ashbrook

Human-Centred Computing Research Section, Copenhagen University

UNIVERSITY OF COPENHAGEN

Introduction

Online how-to videos are popular for learning physical tasks, but current videos lack voice interaction support.

Voice interaction can enhance the learning experience by allowing users to stay focused and hands-on while navigating video content.

Augmented voice playback control features, such as audio icons, have the potential to enhance voice navigation in instructional videos.

Research Questions
RQ1: Can audio icons improve video navigation in how-to videos?
RQ2: What are the challenges and opportunities for designing voice navigation in instructional videos?

Methods

Experimental Design
Utilized a WoZ between-participant design

Independent variable
Augmented navigational audio icon features

Measures
Objective measures: Task completion time, number, and types of commands
Post-task questionnaires: Usability (SUS) and cognitive workload (NASA-TLX)
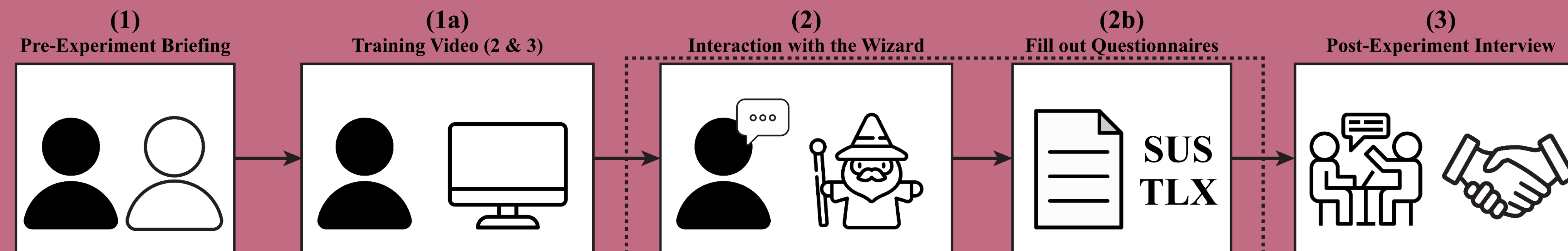Semi-structured interviews: Feedback and suggestions for improvement

Participants
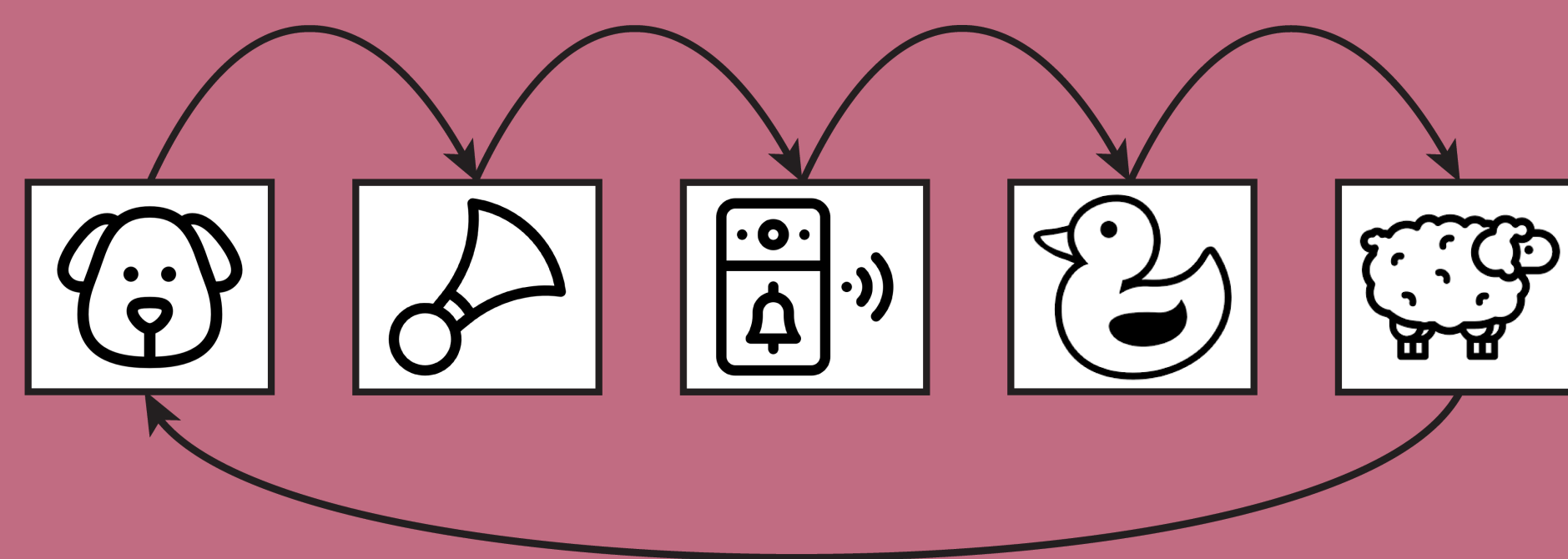24 participants (15 male, 9 female)

Statistical analysis
One-way ANOVA followed by Tukey's HSD post hoc test ($\alpha = 0.05$)

Study Overview

(1) Pre-Experiment Briefing — (1a) Training Video (2 & 3) — (2) Interaction with the Wizard — (2b) Fill out Questionnaires (SUS TLX) — (3) Post-Experiment Interview
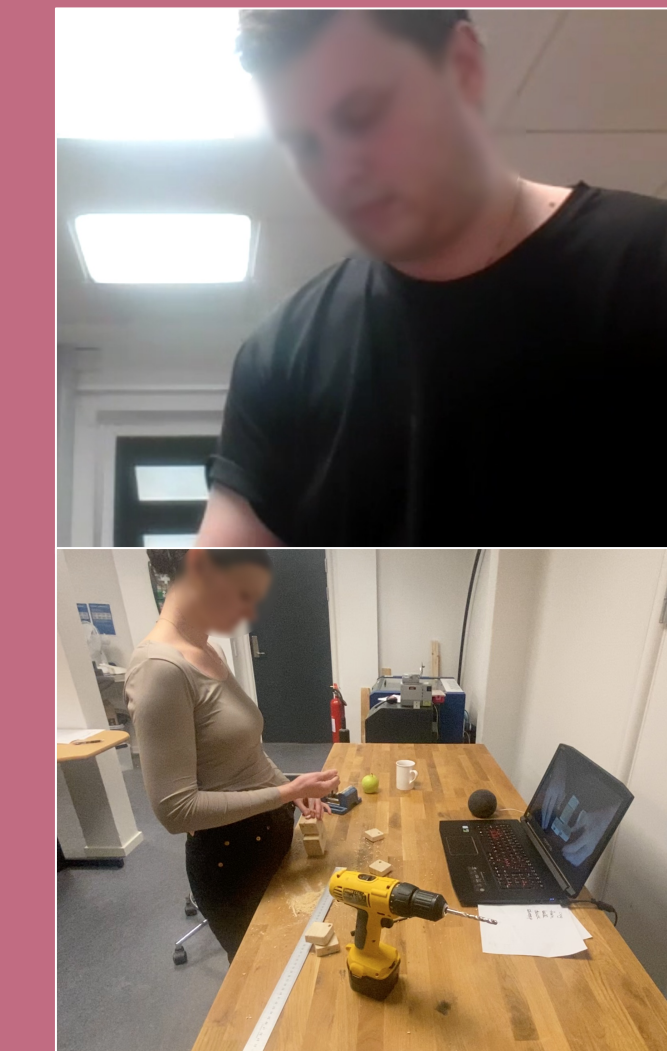
Audio Icons

- Icons appeared 19 times in the how-to video
- Sequence of five recurring sounds: dog bark, car horn, doorbell, duck quack, sheep bleat
- Each audio icon associated with multiple action points

Study Conditions

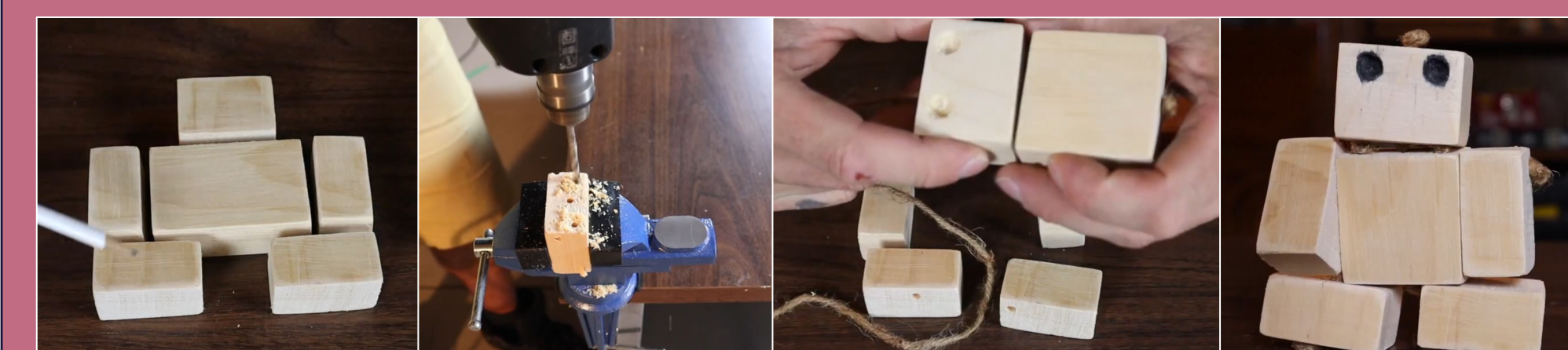|	Study Conditions
Condition 1	Baseline condition with TV remote-like voice control features
Condition 2	Five-audio icon sequence played in conjunction with key points-of-action in the video
Condition 3	Audio icons played at fixed 30-second intervals, serving as landmarks throughout the video

Study Protocol

- Participants were tasked with assembling a wooden robot with a how-to video while using a voice-controlled navigation system that responded to time- and content-based voice commands.
- Performance, usability, and perceived workload were assessed using questionnaires and interviews to evaluate participants' interaction with the video tutorial using voice commands.


Experiment Steps

Contributions

- User study on auditory cues as feedback mechanisms for highlighting events in voice-based navigation.
- Prototype of a hybrid voice interaction system combining temporal and content-based referencing.
- Future research and design opportunities for voice-based non-linear video control.

Results and Conclusion

- Significant difference in usability (SUS) between audio icons at action points vs. baseline condition.
- Participants reported having a positive experience with voice-based playback control, finding it flexible and helpful.
- Audio icons improved voice navigation of how-to videos, resulting in faster task completion, fewer voice commands, and higher SUS scores.
- Participants expressed that the abstraction of meaning between the audio icons and video content hindered effective utilization.
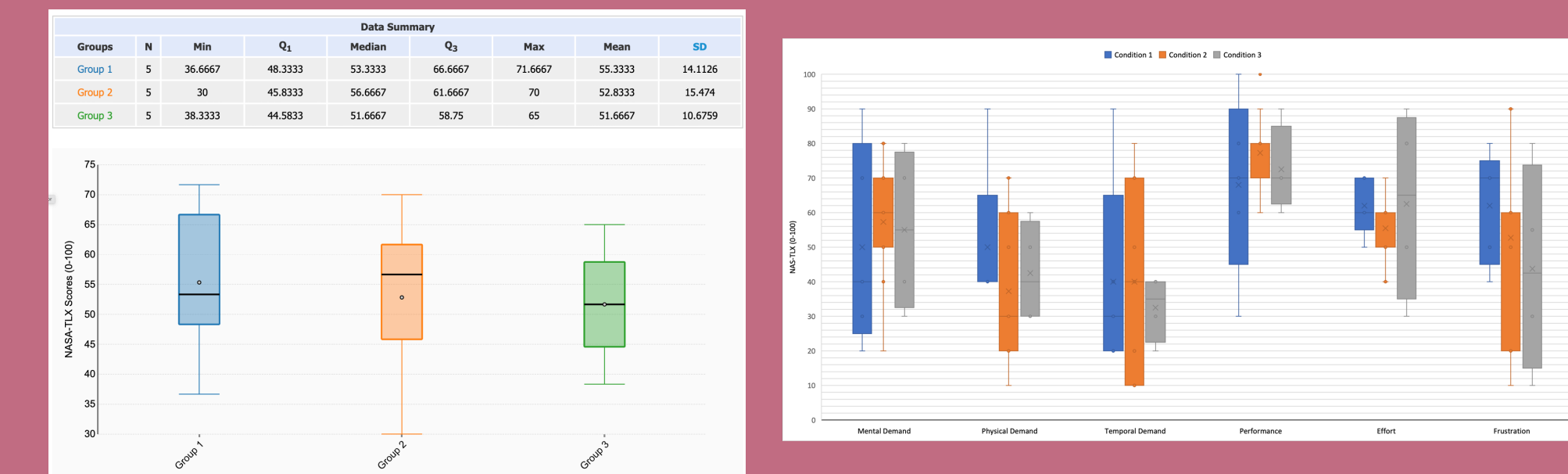
	Average # of Commands	Pause	Play	Rewind	Skip Forward	Time-Specific	Content-Specific	Sound-Specific
Condition 1	40.60	37.93%	40.39%	8.87%	1.48%	11.33%	-	-
Condition 2	30.18	34.94%	38.25%	6.33%	0.90%	7.83%	6.02%	5.72%
Condition 3	28.00	34.82%	38.39%	4.46%	1.79%	17.86%	0.00%	2.68%

Table 4: Average number of commands for each condition and the distribution of command items.