Effects of Visual Modality on Conversations with Interactive Digital Testimonies

Preparing for the Post-Witness Era

Daniel Kolb
daniel.kob@lrz.de
Leibniz Supercomputing Centre
Garching near Munich, Germany

Simona Maiolo*
Ludwig-Maximilians-Universität München
Munich, Germany

Patricia Maier*
Ludwig-Maximilians-Universität München
Munich, Germany

Fabio Genz
Ludwig-Maximilians-Universität München
Munich, Germany

Simone Müller
Leibniz Supercomputing Centre
Garching near Munich, Germany

Dieter Kranzlmüller
Ludwig-Maximilians-Universität München
Munich, Germany

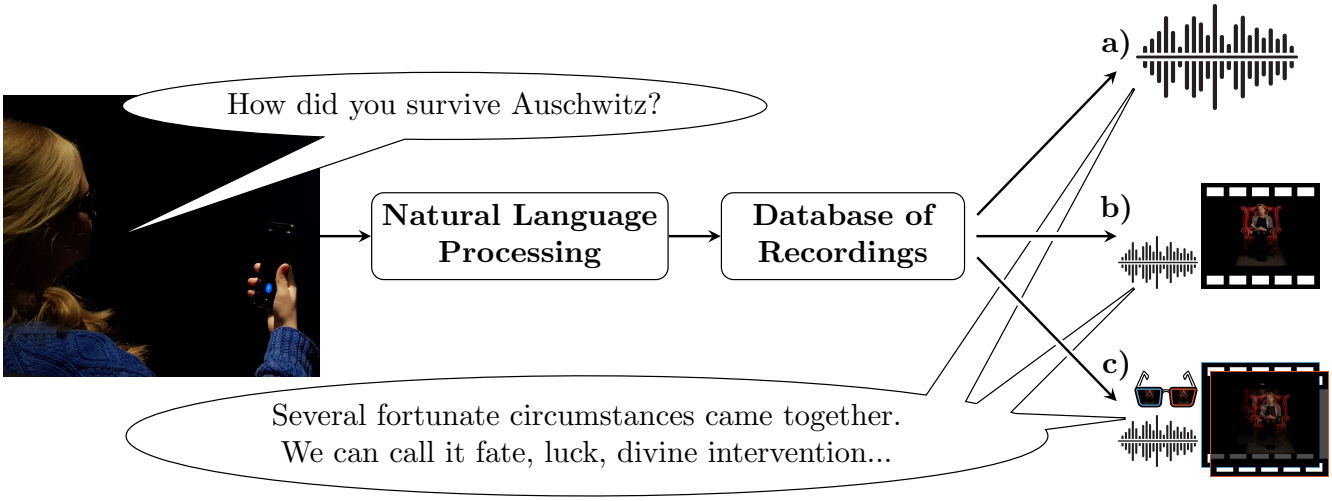


Figure 1: Interactive Digital Testimonies recreate conversations with contemporary witnesses. After processing verbal natural language user input, a matching recording of the response from the original human witness is identified and displayed. Current implementations can present the digital contemporary witnesses a) audio-only, b) in audio-visual 2D, or c) in audio-visual stereoscopic 3D. We investigated how these three output modalities affect users differently.

ABSTRACT

Interactive Digital Testimonies (IDTs) allow users to learn virtually about the life stories of contemporary witnesses as recounted by the witnesses themselves. Although several IDTs have been created in recent years, there is little empirical research on their effects on users. We investigated how different levels of visual modality (audio-only, audio-visual 2D, audio-visual stereoscopic 3D) affect user perception by conducting two separate mixed-methods studies: A 2×2 between-subjects study comparing audio-only with audio-visual 2D in in-person and online settings (n = 82) and a within-subjects study comparing audio-visual 2D with audio-visual stereoscopic 3D (n = 51). We found that audio-visual 2D improves user experience, immersion, and perceived authenticity over audio-only versions. Audio-visual 3D IDTs are more authentic and immersive than audio-visual 2D IDTs, however, this is diminished by a less comfortable interaction. Our findings broaden empirical research on user perception of realistic Embodied Conversational Agents and help guide future thanatosensitive designs.

*Both authors contributed equally to the paper.

CHI '25, April 26-May 1, 2025, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04…$15.00
https://doi.org/10.1145/3706598.3713111

CCS CONCEPTS

• **Human-centered computing → User studies**; *Displays and imagers*; *Virtual reality*; Natural language interfaces; • **Applied computing → Interactive learning environments**.

KEYWORDS

Modality; Presence; Immersion; Learning; Embodied Conversational Agent; Interactive Digital Testimony; Oral History

ACM Reference Format:

Daniel Kolb, Simona Maiolo, Patricia Maier, Fabio Genz, Simone Müller, and Dieter Kranzlmüller. 2025. Effects of Visual Modality on Conversations with Interactive Digital Testimonies: Preparing for the Post-Witness Era. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3706598.3713111

1 INTRODUCTION

Conversations with contemporary witnesses represent an important part of history lessons [13, 16, 105] and citizenship education [14]. Personal experience reports make historical events emotionally comprehensible, tangible [86], and help survivors cope with the trauma they experienced [15, 38, 107]. Unfortunately, contemporary witnesses to an increasing number of historical events, such as Holocaust survivors, are gradually fading away, becoming too weak for long-distance travel or the draining and often emotionally difficult survivor talks. One approach to true-to-life preservation of personal encounters is the virtually simulated dialog by means of Interactive Digital Testimonies (IDTs).

IDTs are Conversational Agents (CAs) using exclusively prerecorded responses. When combined with visual recordings of the witness, they become Embodied Conversational Agents (ECAs). As such, IDTs build on scientific and technological advances in Natural Language Processing (NLP) as well as recording techniques to create interactive virtual twins of contemporary witnesses. The designful absence of AI-generated content prevents IDTs from hallucinating and inventing false statements about the Holocaust [58]. They differ from linear video recordings of interviews with contemporary witnesses, like the Shoah Visual History Archive [97], by giving users control over the selection and sequence of topics. Current implementations support the output modalities audio-only, audio-visual 2D, and audio-visual stereoscopic 3D [7, 40]. The prevailing visual displays range from life-sized partially immersive, projection-based virtual reality systems [5, 71] at designated institutions to consumer-grade hardware for website-embedded IDTs [9]. Unlike fully immersive virtual reality systems using head-mounted displays, these displays readily provide shared experiences for groups of users.

By combining historical contextualization, perspective-taking, and affective connections, IDTs can evoke historical empathy in users. This both cognitive and affective engagement with historical individuals leads to a deeper and more nuanced understanding of history [34, 45]. Especially interactions and conversations with contemporary witnesses can promote strong emotional connections, which increase the quality and enjoyment of learning [14] and prevent presentism [47]. Consequently, genuine-appearing emotive IDTs can foster history and citizenship competences by continually providing opportunities for affective engagement.

Although several IDTs have been developed for different types of displays in recent years, many effects of the chosen modalities on users remain unclear. Since the production process is complex, costly, time-consuming, as well as taxing for contemporary witnesses [98], later amendments, e.g., adding modalities or answers, are rarely an option. With usually only one attempt at capturing a witness' story as faithfully, engaging, and future-proof as possible, planning and implementing IDTs requires meaningful knowledge of the consequences of available design choices. Both, current IDTs as well as future concepts for IDTs and realistic ECAs benefit from thorough evaluations of present-day designs.

To bridge these research gaps, we conducted two distinct user studies. Both studies used two IDTs of Holocaust survivors Eva Umlauf and Abba Naor, created by the "LediZ" (transl. "Learning with digital testimonies") project [12]. Since authenticity and fidelity are key properties of such testimonies, we investigated how differences in visual modality (audio-only, audio-visual 2D, audio-visual 3D) affect user experience, including presence and emotions. In this paper, we use the term 3D to refer to stereoscopic 3D. We employed 2×2 between-subjects design in a study comparing audio-only with audio-visual 2D, in-person as well as online ($n = 82$). In a separate, second study we used within-subjects design to contrast audio-visual 2D with audio-visual 3D ($n = 51$). We used a mixed-methods approach to gather quantitative as well as qualitative data in both studies.

Our paper contains three major contributions:

- Our results extend research on the benefits of embodiment for CAs, by supplying novel data on agents using authentic prerecorded responses instead of synthetically generated replies.
- We detail the advantages and disadvantages of interactions with 3D audio-visual IDTs in comparison to 2D audio-visual IDTs.
- We provide design recommendations for current and future approaches for lifelike posthumous CAs in general and IDTs in particular.

Our results show that audio-visual 2D representations are more immersive, authentic, and pleasant for users than audio-only representations. Audio-visual 3D IDTs are perceived as more authentic and engaging, however, advantages over experiences with audio-visual 2D IDTs are limited by the discomfort caused by 3D glasses.

Our work contributes to fulfilling the research objectives outlined during the CHI'24 workshop on "AI and the Afterlife"[1] [22] as well as in a recent UNESCO report on the potential risks and benefits of AI and digital testimonies in Holocaust education [58]. With the era of the Holocaust witness [113] coming to an end, our findings provide a valuable empirical basis for thanatosensitive [63] design decisions by researchers and educators seeking to preserve conversations with contemporary witnesses for future generations of learners [46].

[1] https://sites.google.com/view/ai-and-the-afterlife-workshop/

2 RELATED WORK

We use Slater's definition of presence as the feeling of "being there" in a virtual environment [95] and immersion as a system property. Co-presence describes the sense of "being together" due to sensory awareness of another co-located actor [17, 41]. Social presence, which builds on co-presence, is the perceived quality of a system to evoke intimacy and immediacy during an interaction [41]. This affects the ability and willingness of users to connect and engage with the virtual actor [17]. The Theory of Interactive Media Effects [99] argues that interactivity, usability, modality, realism, and social presence impact engagement and appreciation of an interactive system. The chosen display modality has consequences for the credibility and likeability of virtual humans as well as the knowledge gained by users. A more immersive system can increase emotional user reactions [109] or induce detrimental side-effects like physical discomfort [116].

2.1 Realism of human-like ECAs

ECAs build on the Computers-Are-Social-Actors (CASA) [73] paradigm: Users interact with computers exhibiting human-like behavior as they would with real humans, even when the users are aware that they are communicating with machines. This makes ECAs powerful interfaces, as users do not have to learn new communication techniques for the face-to-face conversation. Building on simple interactions, they are well-suited to navigate survivor testimonies as well as to strengthen user involvement [37].

The realism of the exhibited human-like behavior influences interaction quality. Multi-modality in in- and output, fidelity of presentation, personality, and overall consistency in behavior are requirements for believable interactive virtual humans [43]. Human realism of embodied agents affects perceived presence and involvement, which influences enjoyment and trustworthiness of the conversation [2, 85]. The choice of display of ECAs needs careful consideration, as it can lead to a credibility loss of the presented information. While agents should be expressive and show distinct emotions, inconsistent [108] or creepy [81] behaviors can have detrimental effects. ECAs with not quite realistic appearances fall victim to the Uncanny Valley Effect [57, 69, 84, 102], which causes feelings of revulsion and aversion as well as distrust [75]. The same effect occurs when body movement and speech are not aligned [101] or the voice does not fit to the visual representation of the ECA [66].

The way an ECA sounds and speaks also contains CASA-relevant social cues [36]. While the effects of the visual appearance of ECAs are well substantiated, research on human-like voices is inconclusive [118]. Previous studies have shown that vocal pitch [33], emotional tone [70], and gender of voice [117] influence the way users perceive and interact with CAs. Similar to the Uncanny Valley, user acceptance decreases if the realism of the voice output does not align with the system capabilities [68]. Voice design of CAs should match their affordances to raise appropriate expectations in users [67]. However, recent findings show that humans prefer realistic voices over synthesized robotic ones [53]. The perceived eeriness decreases with higher degrees of human-likeness, which challenges the existence of an auditory Uncanny Valley. Further studies show that, compared to synthetic voices, human voices cause stronger feelings of trust, social presence [28], and intimacy [82]. While aligning the realism of audio and video of human-like ECAs is beneficial [48, 72], using the respectively highest quality in each modality can result in the highest overall acceptance and trustworthiness [77].

Educational ECAs featuring human voice and human-like behavior cause deeper learning [65]. Hence, ECAs for educational contexts need to appear and act like humans, while avoiding the Uncanny Valley Effect. A study on virtual exhibition guides found that a higher degree of visual realism leads to increased feelings of co-presence [87]. Still, the participants preferred the audio-only representation over abstract and realistic virtual guides, which were perceived as eerie. This also replicated prior findings that voice-only displays can be more successful at evoking emotions and co-presence than abstract visual representations with the same vocal capabilities [8]. Consequently, under some circumstances, non-embodied CAs can rival or even surpass ECAs with regard to user experience and co-presence. Although users expect human-like behavior of educational ECAs [4], realistic behavior can have varying effects on learning, depending on the type of material [80]. Since witness testimonies combine factual and conceptual information, we expect this impact to be comparatively minor. Additionally, the admiration of the person portrayed by the ECA can improve interest and evoke more positive emotions [78].

Previous empirical studies mainly deal with synthetic virtual humans or exposed their shortcomings (see also Table 1). We encountered increased interest in HCI research on the user perception of visually rich ECAs [32]. The results of previous studies on the effects of realistic visual embodiment on perceived social presence [76], co-presence [3], and emotional connection [55] during virtual interactions, however, did not cover ECAs using non-synthetic photorealistic videos as their representation.

2.2 Interactive Digital Testimonies

An IDT simulates conversations with contemporary witnesses by combining a database of prerecorded responses, as given by the original human witness, with a conversational user interface (see Figure 2). They are a subcategory of ECAs, with the restriction that neither the embodiment of the agents nor their replies are synthesized. This aims to create immersive and persuasive audio-visual CAs, without the need for highly realistic computer-generated images, voices, or social cues. IDTs thereby avoid the risks of the Uncanny Valley. However, the exclusion of synthetic replies and the finite number of prerecorded responses limit conversational content and adaptability[2].

Early implementations and predecessors of IDTs date back as far as 1990: "Ask the President" allowed visitors of the Nixon presidential library to choose from more than 280 preselected questions on a touch screen [27]. Upon touching a question, a recording of Richard Nixon's answer was displayed. The recordings were nonuniform due to differences in sources and settings. Visitors were not able to formulate their own questions. The "August system" was a computer-generated recreation of 19th-century author August Strindberg [44]. Users were able to verbally ask the system about the life of the author, among other things.

[2]Additional detailed background on IDT design and production can be found in [10, 56, 98, 103].

Table 1: Overview of related previous studies on human-like CAs and ECAs. The numbers in column n signify the participants in the respective studies.

Modalities	Year	n	Measured Factors (Excerpt)	Output Source	Results	Ref.
Audio-only	1994	180	Personality, Performance	Recordings	Users apply social rules to interactions with computers.	[73]
	2019	30	Disclosure, Closeness	Synthetic	Gender of CA's voice influences self-disclosure.	[117]
	2019	640	Social Presence, Trust	Synthetic	Human voice creates stronger social presence, trust, and behavioral intentions than synthetic voice.	[28]
	2020	95	Human-likeness, Eeriness	Synthetic and Recordings	Human-like voices are perceived as more likable and less eerie than robotic ones.	[53]
Visual-only 2D	2000	40	Likeability, Usefulness	Synthetic	Consistency of verbal and non-verbal cues is preferred.	[48]
	2006	45	Human-likeness, Familiarity, Eeriness	Synthetic and Recordings	Human-likeness increases from humanoids to androids to humans.	[57]
	2023	149	Trust, Uncanniness	Synthetic and Recordings	Virtual influencers appear less trustworthy and human, eliciting less intention to follow recommendations.	[75]
Audio-visual 2D	1999	36	Impression Management, Disclosure, Comfort	Synthetic and Recordings	Consistency across modalities leads to increased impression management, self-disclosure, and comfort.	[72]
	2006	30	Disclosure, Co-presence, Emotion Detection	Synthetic and Recordings	Avatar realism increases co-presence and reduces self-disclosure.	[8]
	2011	129	Human-likeness, Familiarity	Synthetic and Recordings	High-fidelity synthetic human-like characters are especially eerie if their emotional expressiveness is limited.	[102]
	2011	48	Human-likeness, Eeriness, Warmth	Synthetic and Recordings	Mismatching realism in both voice and face elicits feelings of eeriness.	[66]
	2012	172	Emotional state, Empathy	Synthetic	Emotional facial expressions and the tone of voice influence users' feelings.	[70]
	2013	88	Trust	Synthetic	Lower vocal pitch and smiling increase trust in ECAs.	[33]
	2015	113	Human-likeness, Familiarity	Synthetic and Recordings	Asynchrony between lip movement and speech increases the Uncanny Valley Effect.	[101]
	2021	108	Virtual Intimacy, Comprehension, Duration	Synthetic	Voice output (instead of text) and interaction duration increase perceived intimacy.	[82]
	2022	305	Trust, Acceptance, Credibility	Synthetic and Recordings	Natural animation and human-like voice enhance trust, acceptance, and credibility.	[77]
	2022	134	Motivation, Emotion, Likeability, Learning	Synthetic	Likeability of and familiarity with the displayed character increases motivation to learn and positive feelings.	[78]
Audio-visual 3D	2019	20	Co-presence, Human-likeness, Comprehension	Synthetic and Recordings	Visual guides enhance co-presence, while audio guides minimize distraction and improve realism.	[87]
	2021	161	Learning, Enjoyment, Uncanniness, Presence, Cognitive Load	Synthetic	Incongruent realism of ECA appearance and behavior leads to better learning.	[80]

The responses used facial animation and synthesized or manually preprocessed answers. "Synthetic Interviews" enabled users to talk with famous personas by means of speech recognition and film recordings of actors answering in-character [62]. The answer sets included fallback responses which were displayed if the system found no suitable video for a given input. To continue the illusion, periods between responses were filled with videos showing the actor idling in-character. A follow-up project, "Ben Franklin's Ghost", utilized the same concept and Pepper's ghost illusion [50]. User input was limited to 160 preselected questions or keyword-based typing. The responses of both, the August system and the Synthetic Interviews, were partially fictitious. Not all phrases were direct quotations and the way they were presented, including any social cues, were recreations.

The most prominent current use case for IDTs is in Holocaust education. Considerate implementations of interactive digital media can strengthen users' engagement with survivor stories [21]. The first IDT of a Holocaust survivor was developed by the USC Shoah Foundation in 2014 as part of its Dimensions in Testimony project [103]. At the time of writing, they have created more than 50 IDTs, covering nine languages [106]. The Forever Project undertakes a similar approach, producing IDTs for the UK National Holocaust Centre and Museum [56]. The German team LediZ developed two German-speaking IDTs of Holocaust survivors for use in schools and museums [12]. The IDTs of these three projects share a number of design properties: Each testimony features the witness themself talking about their own life story. All response videos were recorded in 3D specifically for use in IDTs. However, the utilized output modalities vary due to the diverse technical circumstances of sites of operation. All videos put the visual focus on the witness, with a black background and minimal furniture visible. Each IDT contains audio-visual content for more than 1000 prompts [5, 52, 56, 103], including an introductory witness account, an idle loop between questions, and neutral fallback responses if no matching response can be displayed.

While non-interactive digital testimonies, such as video testimonies, elicit interest in students, they offer limited immersion due to the lack of interactivity [23]. The current design of IDTs is quasi-interactive [83], as using only prerecorded replies makes it impossible for the CA to reference user-defined verbal input or prior exchanges. While the IDTs were recorded in stereoscopic 3D, the choice of visual display (e.g., 2D, Pepper's ghost, 3D) varies by implementation and location.

2.3 Prior evaluations of Interactive Digital Testimonies

To date, but a small number of empirical studies of IDTs have been conducted and published. Only a subset addressed user experiences and perceptions directly, with the remainder discussing the topic tangentially or referencing internal evaluations, which have not been made available to the public. Nevertheless, these studies highlighted the distinct need for further empirical research on user interactions with IDTs in general as well as on IDT modalities in particular.

We identified two empirical studies of the 3D IDT of Abba Naor by the LediZ project. Both constituted explorative, cursory examinations of user experiences using descriptive statistical analyses. The participants in the first study (n = 46) found interacting with the IDT easy and emotive, but shortcomings of its conversational ability diminished the user experience [52]. It also raised doubts about the importance and impact of displaying the IDT in 3D. The participants in the second study (n = 74) reported developing an emotional connection as well as perceiving the 3D IDT as immersive and like a human [11]. However, they felt that displaying the IDT in 3D was only of minor importance for the overall impact of the witness' stories.

Earlier IDT research investigated the requirements for the finite pool of pre-recorded answers that constitute the IDT output. With approximately 2000 available responses, Dimensions in Testimony's IDT of Pinchas Gutter contained no suitable response to 3.3% of user inputs [6]. The underlying sample of 1667 user questions was gathered over three days and represented 426 distinct statements. However, whether a response would fit to a given question was not decided by the users themselves, but instead by two project-employed annotators. The study showed that a limited set of responses can hypothetically cover a major percentage of inputs due to a large overlap between user questions. The required qualities of a fully automated matching system as well as actual user perceptions were not within the scope of the study. A subsequent between-subjects study compared the effects of this IDT (n = 25), displayed in 2D, with interacting with a real Holocaust survivor (n = 28) [103]. It found that user engagement with the IDT can exceed 50 minutes and that a larger percentage of users could connect with the story of the IDT than with the live survivor. However, the study contained no inferential statistical analysis and lacked numerous details regarding its instruments and procedure.

A case study on digital interactive displays for dialogic remembering identified that the respective museum professionals intended to elicit feelings in co-presence and empathy in users [5]. The list of investigated interactive displays included the 3D IDTs at the National Holocaust Centre & Museum Nottingham, which were provided by the Forever Project. Since the study did not gather or include data on user interactions with IDTs, it did not ascertain whether the intended effects were actually evoked in museum visitors. An evaluation of these 3D IDTs found a high level of average user satisfaction and that 81.6% of answers returned were relevant to the user's question [56]. However, the examined dataset contained only 42 question-answer pairs and omitted several method details, including the sampling method as well as the size and composition of the group of study participants. The evaluation also does not address whether the identified accuracy value is sufficient to convincingly simulate a conversation.

3 METHOD

We split our investigation into two separate studies, each focusing on two modality conditions. This decision was based on having but two comparable IDTs at our disposal as well as the desire to limit participants' emotional and physical attrition. We therefore conducted two individual studies and prioritized internal validity within each study over external validity between both studies.

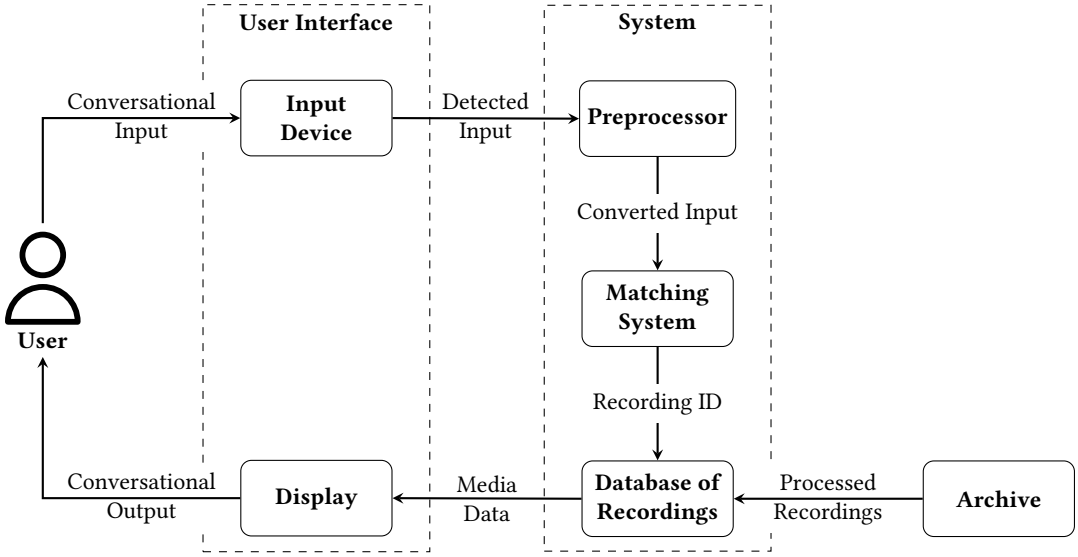


Figure 2: Conceptual model of IDT components and their associated data flows. The components Input Device and Display represent the User Interface, while the Preprocessor, Matching System, and Database of Recordings constitute the corresponding System tasked with the actual processing of the user input. The Archive is used to initialize and populate the Database of Recordings.

(a) Audio-only online **(b) Online audio-visual 2D IDT of Eva Umlauf** **(c) Audio-visual 3D IDT of Abba Naor in-person**

Figure 3: Examples of the visual appearance of IDTs as used in our studies. The white and blue buttons were used when verbally inputting questions.

Study 1 used a between-subjects design to compare audio-only IDTs with audio-visual 2D IDTs. To accommodate participants' needs and increase accessibility, we offered both in-person and socially distanced online participation, which utilized website-embedded versions of the same IDTs. This also allowed us to investigate how different settings affected users. Study 2 used a within-subjects design to contrast audio-visual 2D IDTs with audio-visual 3D IDTs. As suitable 3D displays are uncommon in private households, we required in-person participation for both conditions.

Both studies featured modality as independent variable and used the same two IDTs of Holocaust survivors Eva Umlauf and Abba Naor (see Figure 3) of the LediZ project [12]. With IDTs of two different witnesses, we reduced the influence of individual characteristics like gender [59], personality, or narrative. Both study designs were individually reviewed and approved by our institutions' ethics committee as well as the data protection officers.

We chose the durations of interactions with the IDTs during our studies to be consistent with the way IDTs are offered at educational sites [40], which, in turn, aims to emulate real encounters with witnesses [12]: Each interaction with an IDT lasted up to an hour and included an introductory testimony followed by the opportunity to ask questions. Additionally, we did not provide participants with any specific objectives, tasks, or guardrails and instead allowed the participants to shape the interactions according to their own curiosity and interests. Both align with the durations and designs

of prior user studies on IDTs [11, 52, 103], thereby keeping our study conditions as close to real-world use as possible as well as facilitating the comparability of our findings.

All in-person studies took place in the same cinema-like room which displayed the digital contemporary witnesses in life-size. The Infinity Wall-like screen [29] constituted a partially immersive, virtual reality system [71]. At most six of the 21 available seats were used at the same time due to sanitary restrictions. This helped to maintain the group dynamics of real-life interactions, such as working out questions together or discussing answers, but also uneven use of the input device, e.g., due to shyness. In each session, the light in the room was dimmed regardless of modality. The video data had a resolution of 1920 × 1080 pixels and a 25 Hz frame rate per eye. The projectors were capable of reaching a brightness of 21 000 ANSI Lumen. The 3D display required users to wear polarized glasses to experience the 3D effect. Each IDT contained a twelve-minute introduction by the digital witness telling their story. We provided participants with a preconfigured smartphone as an input device. By pressing and holding down on the touch screen, they were able to ask the IDT verbal questions. All studies used the proprietary cloud-based service Google Dialogflow[3] to analyze the intent of the voice input and select the most fitting prerecorded response, which was then displayed. The corresponding agents were trained in three phases:

- Initialization
- Non-public test interactions
- Public use

The first phase consisted of initializing the training data set with the original interview questions as well as systematic, semantic, or syntactic variations thereof. This enabled first, non-public test interactions with the IDT in the second phase, which refined the matching system and improved its initially low accuracy. The training process implemented supervised learning, with each question-answer-match being manually reviewed and validated or, in the case of deficient classifications, rectified. This review process requires the detailed logging of all user interactions. Both, the systematic variations and supervised learning, presuppose deep topical knowledge to not introduce question-answer-matches that decontextualize or deviate from the original semantic content of the pre-recorded answers [98]. The third phase of training the NLPs is a lasting continuation of the second phase, without being limited to specific users only. The use of centralized NLP systems allowed consistent and simultaneous training and adjustments for all study settings and sites of operation. The application utilized during in-person sessions accessed the locally-stored response files, while the Wowza streaming engine 4.8.5[4] provided the media data via HTTP live streaming to the website-embedded versions. We conducted pilot tests to verify the feasibility of our studies.

Our qualitative evaluations were informed by growing up in a country with a pronounced focus on Holocaust education. Day-to-day encounters with memorials to the suffering caused by national socialism further instilled us with sensitivity to historical responsibility and the dangers of historical negationism. Family ties and friendships with people who experienced and lived through World War II showed us the lasting emotional impact on survivors and the importance of their credibility. These character traits assisted us in empathizing as well as building rapport with study participants and influenced our data coding.

[3]https://cloud.google.com/dialogflow/quotas
[4]https://www.wowza.com/docs

3.1 Study 1: Audio-only vs. Audio-visual 2D × In-person vs. Online

In the first study, we evaluated how audio-only IDTs differ from audio-visual 2D IDTs with regard to user experience, perceived presence, and emotion, as well as accessibility. We utilized in-person and online participation since both IDT modalities can readily be used on-site at educational institutions as well as at home during distance learning.

3.1.1 Participants. A total of 82 participants took part in either the in-person (n = 40) or online study (n = 42). Of these, 54 (66%) identified as female, 28 (34%) as male, and none as non-binary. Their age ranged from 19 to 80 ($M = 34.27, SD = 14.37$). Seven participants (9%) had interacted with at least one IDT prior to this study. We provided assistance if requested, e.g., when reading or filling in the questionnaires. All participants indicated that the IDT's language was their native language. Only one stated that they spoke dialect. The group interacting with the audio-only mode consisted of 36 (44%) participants, 19 (23%) online and 17 (21%) in-person. The remaining 46 (56%) participants experienced the audio-visual 2D mode, 23 (28%) online and 23 (28%) in-person. We recruited participants via social media channels and university mailing lists. We incentivized participation in the in-person study with a 10€ voucher or extra university credit. The participants of the online study could enter a raffle for one of ten 20€ vouchers or receive extra university credit. The compensation system was directed by our institutions' guidelines. The different rates corresponded to the time and effort required to participate in the study, as determined during our pilot tests. Participation in our study was voluntary and not required by any institution.

3.1.2 Study Design. We conducted a between-subjects study with *Modality* (levels: *Audio-only* and *Audio-visual 2D*) and *Setting* (levels: *In-Person* and *Online*) as independent variables. The participants were randomly assigned to a modality level. While the in-person study offered a life-size screen and ensured that sound and image were free from interference, users in the online study could participate in a socially distanced and location-independent manner. Not having to travel made it easier for people with disabilities to partake in this study. Both ways of participation used the same modalities and provided the same media files. We counterbalanced the distribution of the participants to all conditions and balanced the groups in all conditions to control for potential confounds caused by the individual characteristics of the IDT ($n_{Eva\ Umlauf} = 40$, $n_{Abba\ Naor} = 42$).

3.1.3 Measured Variables. With our questionnaire, we measured six variables from a total of 41 questions (see Table 2). Where necessary, we adapted or rephrased questions of these previously validated questionnaires for the context of IDTs and normalized the scales to five points. The questions on user experience (Efficient, Easy, Exciting, Interesting) originated from the short version of the User Experience Questionnaire (UEQ-S) [89]. We measured how

Table 2: Variables measured in the audio-only vs. audio-visual 2D study. Due to the high proportion of non-normally distributed items [104], we used Greatest Lower Bound (GLB) [115] to calculate the internal consistency reliability.

Section	Variable	Items	Reference	Internal Consistency
Questionnaire	User Experience	4	UEQ-S [89]	0.50
Questionnaire	Presence	12	IPQ [90], WS [114]	0.83
Questionnaire	Pos. Emotions	10	PANAS [112]	0.85
Questionnaire	Neg. Emotions	10	PANAS [112]	0.86
Questionnaire	Accessibility	4	WCAG 2.0 [25]	0.68
Questionnaire	Preference	1	-	-
Interview	Reasoning	7	-	-

present and immersed the participants felt with three items (SP3, INV2, INV4) of the igroup presence questionnaire (IPQ) [90] and nine items (5, 6, 8, 9, 21, 23, 25, 26, 32) of the presence questionnaire by Witmer and Singer (WS) [114]. The Positive and Negative Affect Schedule (PANAS) [112] was used to evaluate their emotional responses. We based the questions regarding accessibility on the four WCAG 2.0 guidelines [25]. We also asked users whether they would have preferred the other modality. Free-text fields allowed users to further detail their feedback and impressions. During the in-person study, we conducted voluntary interviews after the questionnaire to gather subjective reasonings on preference, realism, emotiveness, and accessibility of the modality. We used fully structured interviews with open-ended questions to control their length. With a predictable and short duration, we aimed to increase the willingness of participants to be interviewed or to wait for their turn. While this restricted the depth of the interviews, it allowed us to gather more representative and diverse insights.

3.1.4 Study Procedure. Both study settings, in-person and online, used the same procedure (see Figure 4a) and questionnaire. We conducted 13 in-person sessions, with group sizes ranging from two to six people ($M = 3.08, SD = 1.61$). Online users took part individually and on their own. Each session began with an explanation of our methods of data collection and processing. Participants were free to suspend or discontinue to partake in the study at any time and for any reason. If they gave their consent, they subsequently experienced a digital contemporary witness sharing their story for twelve minutes. The participants then interacted with the digital contemporary witness for up to 30 minutes. The shortest interaction duration was 12 minutes for the audio-only setting with a group of two users and 15 minutes for the 2D display with a group of three users. The introduction and the interaction were presented in the same modality and with the same IDT. After the interaction each participant filled in the questionnaire. The overall duration of each session was limited to 60 minutes. 31 participants (78%) voluntarily expanded on their feedback in a short interview at the end of an in-person session. Volunteers from the same session were interviewed individually, separately, and successively.

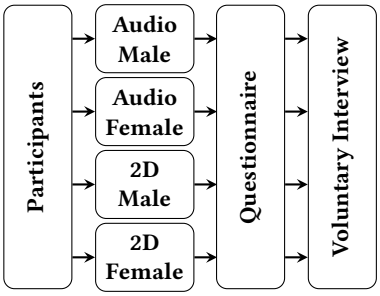
3.2 Study 2: Audio-visual 2D vs. Audio-visual 3D

The second study evaluated how audio-visual 2D IDTs differ from audio-visual 3D IDTs with regard to user experience, perceived presence and emotion, physical discomfort, and user preference.

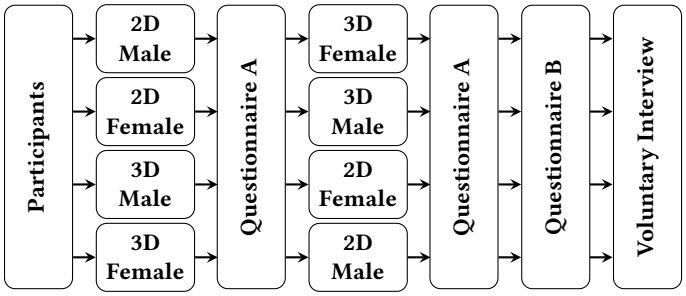
3.2.1 Participants. We recruited 51 participants through university newsletters and social media channels. We offered them either extra university credit for their studies or 20€, as directed by our institutions' guidelines on incentives. The compensation amount was based on the time and effort required to participate in the study, which we determined during our pilot tests. Participation in the study was voluntary and not required by any institution. In our sample, 30 (59%) identified as female, 21 (41%) as male, and none as non-binary. Their age ranged from 19 to 65 ($M = 28.4, SD = 9.3$). Nine participants (18%) had interacted with at least one IDT prior to this study. Two participants (4%) stated that they had a high level of proficiency in the IDT's language, with the remainder (96%) indicating a very high or native level of proficiency.

3.2.2 Study Design. We conducted a between-subjects study with *Modality* as the independent variable. The levels were *Audio-visual 2D* and *Audio-visual 3D*. All participants interacted with both display modes. We used IDTs of different Holocaust survivors to reduce undesired learning effects for the second interaction. We counterbalanced the random distribution of the participants to the four study sequences.

3.2.3 Measured Variables. With our questionnaires, we measured five variables from 29 questions. The structures and origins of the questionnaires are presented in Table 3. We based our items on previously validated questionnaires, which we adapted or rephrased for use with IDTs and normalized to a five-point scale, where necessary. Questionnaire A surveys the users' perceptions after interacting with each given display mode. Based on the UEQ-S [89], it measures the general user experience (Efficient, Easy, Exciting, Interesting). Since Study 2 featured visual output in both conditions as well as a more immersive display method in one condition, we utilized different targeted questionnaires than in Study 1. With 14 items of the Immersive Experience Questionnaire [49] we evaluated how present and immersed the participants felt during the interaction. We measured their emotional responses during the interaction with six items (CJOA2D, CJOC1B, CAGC1D, CAXM2D, CAXP2D, CSHC3D) from the Achievement Emotions Questionnaire



(a) Study 1 used between-subjects design to compare audio-only and audio-visual 2D IDTs in in-person and online settings.



(b) Study 2 used within-subjects design to compare audio-visual 2D with audio-visual 3D IDTs.

Figure 4: Overview of procedure and design of both studies.

Table 3: Variables measured in the audio-visual 2D vs. audio-visual 3D study. Due to the high proportion of non-normally distributed items [104], we used GLB [115] to calculate the internal consistency reliability.

Section	Variable	Items	Reference	Internal Consistency
Questionnaire A	User Experience	4	UEQ-S [89]	0.77
Questionnaire A	Presence	14	IEQ [49]	0.97
Questionnaire A	Emotions	6	AEQ [79]	0.89
Questionnaire A	Discomfort	4	SSQ [51]	0.75
Questionnaire B	Preference	1	-	-
Interview	Reasoning	6	-	-

(AEQ) [79], due to its reliability and suitability for immersive virtual environments [100]. Since stereoscopic displays can cause physical discomfort, we surveyed the well-being of the participants with four questions (General Discomfort, Fatigue, Eyestrain, Nausea) originating from the Simulator Sickness Questionnaire (SSQ) [51]. After interacting with both modes, Questionnaire B asked users to select their preferred modality. We gained further insights through free-text fields and by conducting voluntary interviews on participants' reasonings for preference, perceived realism, and comfort of use of the modalities. We used fully structured interviews with open-ended questions to build a representative and diverse sample. The predictable and short duration aided the participants in accommodating our interview request.

3.2.4 Study Procedure. The procedure of this study is presented in Figure 4b. We conducted 16 sessions, with group sizes ranging from one to five people (M = 3.19, SD = 1.60). We initiated each session with an explanation of our methods of data collection and processing. This included informing the participants that they were free to suspend or discontinue taking part in the study at any time and for any reason. If they consented to the terms, they were shown an introductory video of a digital contemporary witness recounting their story within twelve minutes. The subsequent interaction with the digital contemporary witness used the same display mode. Each participant filled in Questionnaire A regarding their experience with the first display mode. This was followed by a short break and the analogous procedure for the second IDT in the respectively other display mode and another iteration of Questionnaire A. Afterwards, the participants filled in Questionnaire B. To control variance in time spent interacting with the IDTs, we advised users to ask their last question after 24 minutes. We also ended the interaction if the participants had no more questions for the IDT. The shortest interaction duration was 11 minutes for the 2D display and 12 minutes for the 3D display. Both were sessions with a single participant. The overall duration of each session was limited to 120 minutes. 30 participants (59%) voluntarily expanded on their feedback in a short interview at the end of a session. Volunteers from the same session were interviewed individually, separately, and successively.

4 RESULTS

We analyzed the quantitative and qualitative data of Study 1 and Study 2 individually.

4.1 Audio-only vs. Audio-visual 2D × In-person vs. Online: Quantitative Results

In the following, we list the quantitative results of Study 1. For our analysis, we first tested our sets of measured variables for normal distribution and homogeneity of variance. While we found isolated violations of the assumptions of normality or homogeneity of variance, ANOVA is considered generally robust to both these violations, if the number of participants in each group is approximately equal and not unreasonably small [18, 64]. We consequently carried out factorial ANOVA to identify significant ($p < .05$) impacts of modality and setting, as well as interaction effects. Tukey's HSD test was used for post hoc analyses. Table 4 summarizes the main

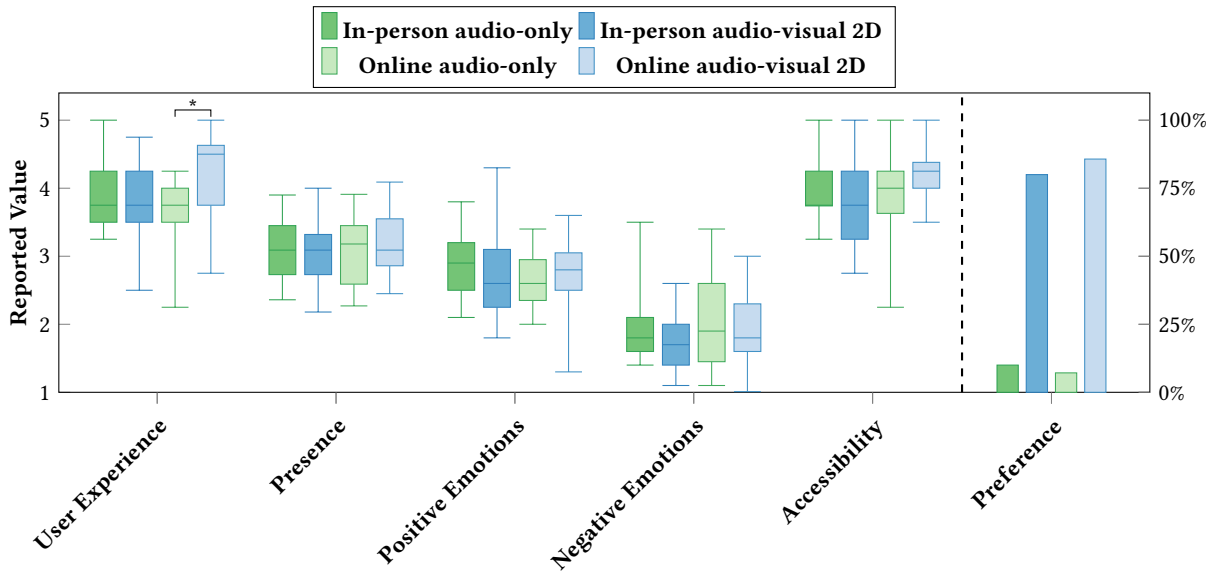


Figure 5: Aggregated measures of the audio-only vs. audio-visual 2D × in-person vs. online study. * indicates a significant (p < .05) pair-wise difference. Separated by a dashed line are the percentages of participants who would have preferred the respectively opposite modality. The totals for preference are less than 100%; the remainder felt indifferent.

findings. The aggregated values and the distribution of preference are shown in Figure 5.

We found a significant interaction effect of modality and setting in overall *User Experience* ($F(1, 78) = 5.65, p = .02$). On the whole, audio-visual 2D IDTs provided online users with a better user experience than audio-only IDTs ($p = .02, d = 0.54$). The other pairs of levels were not significantly different. However, an analysis of the sub-scales showed significant differences in ratings of the item *Easy* for setting ($F(1, 78) = 10.35, p = .002$) as well as an interaction effect ($F(1, 78) = 5.58, p = .02$). In-person participants found talking with IDTs easier than online participants ($p = .002, d = 0.62$). Particularly online audio-only users reported lower ratings than the in-person audio-only group ($p = .001, d = 1.13$) and the in-person audio-visual 2D group ($p = .02, d = 0.80$). Additionally, modality had a large effect on the item *Exciting* ($F(1, 78) = 21.64$), with audio-visual 2D IDTs eliciting more excitement in users than audio-only IDTs ($p < .001, d = 0.86$).

The analysis of differences in perceived *Presence* showed no significance caused by modality ($F(1, 78) = 0.48, p = .49$), setting ($F(1, 78) = 1.55, p = .22$), or their interaction ($F(1, 78) = 0.21, p = .65$). While factorial ANOVA returned significant differences for WS23 ("I asked the interactive digital testimony many questions.", $F(1, 78) = 5.58, p = .02$), the post hoc test found no significant pairs. However, we found that WS25 ("The interactive digital testimony reacted quickly to my questions.") was significantly influenced by setting ($F(1, 78) = 10.72, p = .002$) and modality ($F(1, 78) = 5.72, p = .02$). Online users perceived that the IDTs reacted more quickly than in-person users ($d = 0.65$). Participants who interacted with the audio-visual 2D testimony reported shorter delays between input and output than those who interacted with the audio-only testimony ($d = 0.48$).

Our quantitative analysis showed no significant effects of modality ($F(1, 78) = 0.001, p = .98$), setting ($F(1, 78) = 0.48, p = .49$), or their interaction ($F(1, 78) = 0.49, p = .49$) on positive emotions experienced during conversations with IDTs. Similarly, we found no statistically significant impacts by modality ($F(1, 78) = 1.75, p = .19$), setting ($F(1, 78) = 1.68, p = .20$), or their interaction ($F(1, 78) = 0.90, p = .35$) on overall negative emotions. However, setting caused a difference in how *Distressed* users felt ($F(1, 78) = 6.48, p = .01$). Talking with the digital witnesses about their life stories moved the in-person participants emotionally more strongly than online participants ($p = .01, d = 0.70$).

Overall accessibility was rated similarly high by all groups. Our quantitative analysis found no significant impacts by modality ($F(1, 78) = 0.65, p = .42$), setting ($F(1, 78) = 2.39, p = .13$), or their interaction ($F(1, 78) = 3.51, p = .06$). An analysis of the sub-scales found interaction effects for the items *Understandable UI* ($F(1, 78) = 8.22, p = .09$) and *Operable UI* ($F(1, 78) = 6.48, p = .01$). Online users of the audio-visual 2D IDT found the interface easier to understand than online users of the audio-only IDT ($p = .03, d = 0.73$) as well as the in-person participants interacting with the audio-visual 2D IDT ($p = .01, d = 0.74$). The post hoc test for *Operable UI* found no significant pair-wise differences.

In the audio group, 88% of in-person users and 84% of online users would prefer interacting with an embodied IDT. Overall, 80% of the 2D group would not want to forgo the visual modality.

4.2 Audio-only vs. Audio-visual 2D × In-person vs. Online: Qualitative Results

Our reflexive thematic analysis [19, 20] used both the transcribed interviews as well as the free-text explanations on the questionnaires from all four groups. The individual participants are listed

Table 4: Factorial ANOVA results of the modality × setting study across the levels {audio-only, audio-visual 2D} × {in-person, online}. For this overview, we shortened "audio-visual 2D" to "AV2D". * indicates p < .05, ** signifies p < .01 and * is p < .001.**

Variable	Statistic	Effect size	Levels comparisons (Mean, SD)
User Experience			
Modality × Setting	$F(1, 78) = 5.65^{*}$	$\eta^2 = 0.06$	AV2D+Online (4.14, 0.69) > Audio+Online (3.61, 0.53)
Easy			
Setting	$F(1, 78) = 10.35^{**}$	$\eta^2 = 0.11$	In-Person (4.58, 0.50) > Online (3.95, 1.15)
Modality × Setting	$F(1, 78) = 5.58^{*}$	$\eta^2 = 0.06$	Audio+In-Person (4.76, 0.44) > Audio+Online (3.63, 1.12), AV2D+In-Person (4.43, 0.51) > Audio+Online (3.63, 1.12)
Exciting			
Modality	$F(1, 78) = 21.64^{***}$	$\eta^2 = 0.21$	AV2D (4.59, 0.62) > Audio (3.72, 1.06)
Short Response Time			
Modality	$F(1, 78) = 5.72^{*}$	$\eta^2 = 0.06$	AV2D (4.07, 0.93) > Audio (3.58, 1.00)
Setting	$F(1, 78) = 10.72^{**}$	$\eta^2 = 0.11$	Online (4.17, 0.96) > In-Person (3.52, 0.91)
Distressed			
Setting	$F(1, 78) = 7.19^{**}$	$\eta^2 = 0.08$	In-Person (3.29, 1.09) > Online (2.58, 1.30)
Understandable UI			
Modality × Setting	$F(1, 78) = 8.22^{**}$	$\eta^2 = 0.09$	AV2D+Online (4.48, 0.59) > AV2D+In-Person (4.17, 0.78), AV2D+Online (4.48, 0.59) > Audio+Online (3.89, 0.94)

in Table 6 in Appendix A. The first author was involved with the manual coding and construction of themes from shared meaning-based patterns. These patterns were focused on, but not limited to, experiences related to the output modality. Our inductive coding of primarily semantic meanings built on a predominantly experiential and constructionist interpretation of the data [24]. We created candidate themes and sub-themes after the first coding iteration. Over subsequent repetitions, we reflected on and revised codes as well as themes accordingly. The final revision, which took place ten weeks after the first iteration, resulted in the following themes and sub-themes, which can also be seen in Figure 6:

4.2.1 Tactful visuals promote engagement and emotional connection: Participants across all groups reported enhanced cognitive and affective engagement as well as social presence [93] as a prominent property of visual representations of virtual humans. However, displays featuring visual details besides contemporary witnesses can introduce additional distractions.

Attention anchor: The visual stimuli of audio-visual 2D IDTs immediately and firmly attracted the attention of respective users. By continually occupying more senses, they were able to retain this attention for longer periods of time and further accentuate the contemporary witness. This made it easier for participants in the audio-visual 2D groups to concentrate on the narration and interaction: *"I thought it was very visually appealing, so it was easy to focus on her and for me it's also important to have video and not just audio, because it just helps me focus more" (P37).* Conversely, participants in the audio-only groups were inclined to let their gaze, and thus their attention, wander due to the absence of a predominant visual anchor. For some, this resulted in heightened awareness of and even distractions by their surroundings, while others found the experience tedious: *"Without visual stimulation, just listening for 10 minutes is a bit long" (P55).* Overall, captivating users visually fostered immersive, fulfilling, and memorable interactions.

Non-verbal social cues: Additionally, the visual output added inaudible nuances of the Holocaust survivors' responses, which participants in the audio-only groups missed: *"I just think that it would be better, for example, with visuals, actually. Not so long ago, I had a conversation with a Holocaust survivor, and so much was transmitted through his facial expressions and gestures, especially regarding his descriptions of Auschwitz, but also his personality" (P13).* Perceiving the body language or even tears of the contemporary witnesses made it easier to sense and understand the emotions than from their voices alone. It also further added to the candid and intimate nature of the conversation. Participants who interacted with the audio-visual 2D IDT remarked that they felt emotionally closer and connected more strongly because they were able to also experience these soundless reactions. P26 illustrates this with the difficulty of talking about certain historical events, as exhibited by the IDT of Abba Naor: *"I don't think it was always easy for him, because he always stroked along the edge of the chair. You already noticed that it is very moving and difficult for him, but it comes across somehow more clearly in such a visual case. [...] And that is then just different on an emotional level as if you only hear it" (P26).* Some participants in the audio-only groups experienced difficulties with the conversational flow, as they were uncertain whether silent phases stemmed from the conclusion of a response or from a pause for thought. They consequently felt less connected or immersed and equated the interaction to a phone call instead of a face-to-face conversation. In accordance with the Media Richness Theory (MRT) [30], this also demonstrates that the visual output supplies users not only with auxiliary emotional cues but also functional cues like information about turn-taking [26].

Visual clutter: While visual displays can enhance the quality of the interaction with IDTs, participants across all groups agreed that overcrowded or inappropriate visual content can also detract from the experience. A suitable visual representation should be neither dull nor distracting: *"The simple design (black background,*

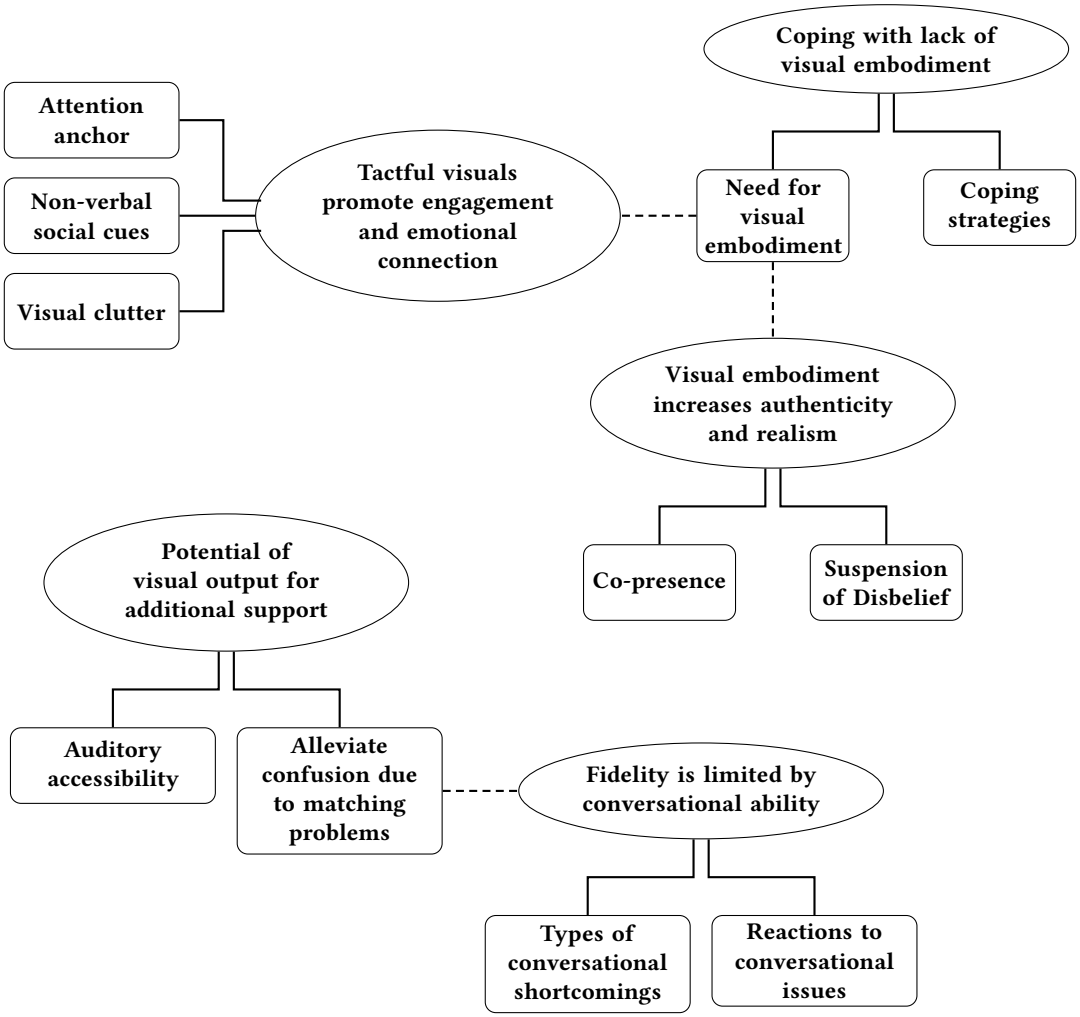


Figure 6: The themes and sub-themes derived from our analysis of interviews and free-text explanations of Study 1.

chair, human) looks intriguing due to the discreetly pompous antique chair and the elegantly dressed contemporary witness but still does not attract too much attention so that the user could be distracted from the essential" (P71). Caution must be taken when adding visual features or details, as these can simultaneously add mental barriers and distance between users and the virtual representation of the witnesses or even detract from the witnesses themselves. Instead of attempting to artificially evoke or amplify emotions in users, it is essential to let contemporary witnesses and their narrations speak for themselves.

4.2.2 Coping with lack of visual embodiment: The interactions with the IDTs coincided with the desire to see the virtual conversational partners. Participants in the audio-only groups employed different strategies in an effort to satisfy this need with varying success.

Need for visual embodiment: Those who interacted with the audio-only IDTs voiced their disappointment over the absence of any visual representation of the respective Holocaust survivor: *"I thought that was a pity because I would have very much liked to see his face, the person behind the voice" (P4).* Besides fulfilling their curiosity, the participants reasoned that they wanted to be able to recognize the contemporary witnesses in other contexts and forms of media. Audio-visual IDTs can thus provide users with a more enduring and resurging impression.

Coping strategies: When facing audio-only IDTs, many participants felt compelled to try to imagine what the respective contemporary witnesses might look like. While some were content with their mental images, others were dissatisfied with the approach: *"You want to see who you're talking to... I couldn't picture the woman, which significantly reduced the quality of the interaction" (P46).* In addition to difficulties when trying to visualize the Holocaust survivors, participants were ultimately aware that the imagined appearances were not their real appearances. Their need consequently remained unfulfilled. To remedy this, some participants chose to look up the contemporary witnesses online while listening to the IDT. Side activities like operating their own smartphones in the in-person setting or using their browser in the online

setting meant diverting their focus and reducing the immersion. The third strategy was using the surrounding facility or devices as visual substitutes. One participant who interacted with the audio-only IDT of Eva Umlauf and had seen the real person before in a different context cautioned: *"If I had never known her face, then it would be hard for me to contextualize. Then that would actually only ever be connected to, I don't know, this room and less to experiencing a person" (P1).* Consequently, no coping strategy was able to fully satisfy the users' needs without detracting from the experience.

4.2.3 Visual embodiment increases authenticity and realism: We found that visual embodiment aids the authenticity of the presented information as well as the perceived realism of the digital witness. Realistic dimensions of the embodiment can raise the authenticity and immersion further. However, no participant remarked that they experienced their respective modality as inherently inauthentic.

Suspension of Disbelief: Across all groups, participants were essentially aware that they were not interacting with a real person. However, facilitated by the lifelike visual embodiment of the contemporary witnesses, users were able to suspend their disbelief. After interacting with the audio-visual 2D IDT of Eva Umlauf, P18 reported: *"I really enjoyed it. Like, there was this chair and I thought at some point, this is actually real. She is really sitting there" (P18).* In addition to consistent movements and reactions reported by both audio-visual groups, participants in the in-person audio-visual 2D setting also highlighted the true-to-life size and proportions of the digital witnesses as causes for the perceived realism. While the auditory output itself was perceived as realistic, most respondents in the audio-only groups missed a display of the original and identifiable human source. They, therefore, had diminished reason not to perceive their conversational partners as computer systems. Additionally, ambiguity in the origin of the IDT's responses could foster doubts in the validity of its content: *"Without video it feels anonymous and one could also claim that what was said was 'fake' or not a recording of a real witness. With video, the whole thing would be even more credible and impressive" (P54).*

Co-presence: Exclusively participants in the in-person audio-visual 2D setting mentioned feeling as if being in the same room [41] with the digital witness. This shows that a visual embodiment is necessary for, but does not ensure, evoking co-presence in sighted users of IDTs. Building on suspension of disbelief and engagement, it benefits from presenting the digital witness in a realistic size and in a calm environment, physical and virtual. In our case, this elicited the imagination of being invited and hosted by the contemporary witness: *"I thought it was somehow beautiful, as well as aesthetic, how she sat there in her armchair and you got the feeling that you are at her home with her" (P40).*

4.2.4 Fidelity is limited by conversational ability: All groups determined conversational flaws as the main cause for disruptions in the perceived fidelity of the IDT. Encountering weaknesses of the CA elicited diverse reflex reactions, which diminished suspension of disbelief and engagement.

Types of conversational shortcomings: We identified three types of conversational issues reported by our participants. The first type was receiving fallback answers if no matching response was recorded and available. The limited size of the pool of answers is inherent to the concept of IDTs and its categorical avoidance of procedurally generated, synthetic responses, which precludes follow-up questions or references to prior exchanges as well. However, this shortcoming aligned most with user expectations and was met with leniency: *"It is easy to use and comprehend. However, the answers understandably cannot adequately cover all questions" (P45).* The second type was the CA failing to understand too complex or deeply nested questions. This prompted participants to consciously or unconsciously adjust and deviate from their accustomed way of phrasing questions: *"I think formulating the questions was more difficult because you tried to make them as simple as possible and couldn't ask spontaneously" (P25).* The third and most salient type was receiving an incorrectly matched, unsuitable answer. Most of the participants' more distinct reactions originated from these matching errors.

Reactions to conversational issues: The effect of the conversational shortcomings on participants was twofold: The issues elicited diverse emotional reactions and reduced the immersion. Incorrectly matched responses were still able to provide participants with interesting information. Some participants felt indifferent and simply accepted the issues as technological restrictions, which nonetheless broke the illusion of talking with a real human being. Amusement or levity, however, caused participants to feel conflicted: *"So as soon as the technology fails and doesn't work, you kind of always react with humor and find it funny, but in that context I don't know if that's okay" (P20).* In most cases, the conversational shortcomings caused frustration or disappointment over the broken suspension of disbelief: *"In the moments when she gave exactly the answer that matched the question. Then it was briefly a 'Wow'! And otherwise you noticed that it's just an artificial situation" (P32).* These feelings were then accompanied by heightened awareness of the surroundings, disrupted engagement, and gaps in the flow of the interaction: *"But you're always pulled out of the interaction a bit and come back to the real world when he can't answer the question. And then you're quickly pulled out of the whole conversation" (P8).* Unmitigated conversational issues reinforced the recognition of IDTs as unalive computer systems, which can lead to potentially undesirable perceptions of and interactions with digital witnesses.

4.2.5 Potential of visual output for additional support: A visual display offers the opportunity to provide users with supplementary information and feedback. This could address accessibility issues as well as improve overall usability.

Auditory accessibility: Irrespective of the type of modality provided during the study, users occasionally experienced difficulties comprehending the spoken verbal output. However, these were tied neither to the audio quality nor the users' hearing abilities or language proficiencies. Along with the dialect and articulation of the human witnesses, the unmodified recordings preserve corresponding issues: *"Due to his accent, I did not understand all the place names" (P82).* Additionally, the narrations contain uncommon words and terms that participants were unfamiliar with. For these instances, optional visual information, like captions or subtitles, can provide additional support. The benefits could extend to native speakers, second-language speakers, hearing-impaired users, as

well as hearing users: *"There are also, for example, Jewish names or Hebrew names that I don't know at all. And I think captions would be relatively helpful, also if one can no longer hear so well" (P3).*

Alleviate confusion due to matching problems: The participants also proposed using the visual output to mitigate conversational issues. Displaying the verbal input, as recognized by the NLP, could inform users about improper use or limitations of the input device. Providing the original question, which was asked during the recording process, of currently displayed responses could add context to incorrect matches: *"It doesn't actually frustrate me when the questions don't fit so perfectly or the answers don't match the questions. But I would like to know the actual question she is responding to" (P1).* Both types of feedback would increase the transparency of the quality of the matching process. However, as outlined in Section 4.2.1 and Section 4.2.3, they could also divert attention from the digital witness as well as emphasize the non-human nature of the conversational partners.

4.3 Audio-visual 2D vs. Audio-visual 3D: Quantitative Results

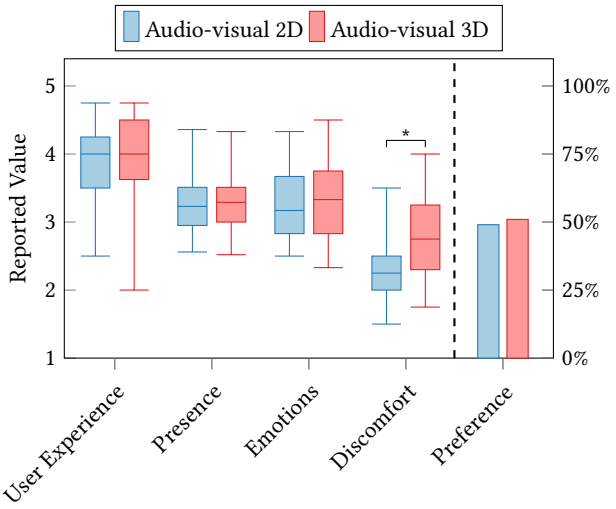


Figure 7: Aggregated measures of the audio-visual 2D vs. audio-visual 3D study. * indicates a significant (p < .05) difference. The total for preference is 100%.

In the following, we list the quantitative and qualitative results of Study 2. For our analysis, we first tested the sets of measured variables for normal distribution. If the Shapiro-Wilk test returned p > 0.05 for both, the 2D and 3D data sets, we used paired t-tests to investigate their mean differences and Cohen's d to measure the effect size. For measurements with non-normal distributions, we used paired Wilcoxon signed-rank test and its corresponding effect size r. Only the data sets for *Presence* and *Emotion* are normally distributed. The aggregated values and the distribution of preference are shown in Figure 7. Table 5 summarizes our main findings.

Our participants reported similar *User Experience* for audio-visual 2D IDTs ($M = 3.87$, $SD = 0.57$) and audio-visual 3D IDTs ($M = 3.92$, $SD = 0.64$). We conducted a paired Wilcoxon signed-rank test, which found no significant effect ($p = .57$, $r = 0.08$). The analysis of the items *Efficient*, *Easy*, *Exciting*, and *Interesting* also showed no significant variance between modalities. Participants rated stimulating features higher than functionality features. Notably, five users found the 3D display less interesting and exciting than the 2D display. One participant, who valued the 3D display mode lower for all user experience items, argued that 2D is sufficient. They claimed that the only noticeable 3D characteristic was the depth of field between the contemporary witness and the back of the chair, which appeared unnatural.

A paired t-test showed no statistically significant change in perceived *Presence* ($t(50) = -1.48$, $p = .14$, *Cohen's* $d = 0.04$) when interacting with a 3D testimony ($M = 3.34$, $SD = 0.49$) instead of a 2D testimony ($M = 3.36$, $SD = 0.44$). However, we found that participants paid more *Attention* ($p = .02$, $r = 0.34$) to the conversation in 3D mode ($M = 2.27$, $SD = 1.34$) than in 2D mode ($M = 1.67$, $SD = 0.79$). Participants remarked that the 3D display conveyed more spatial depth, which made them feel more co-present. Participants who felt more present in the 2D display mode argued that 2D IDTs were less strenuous for their eyes.

Using paired t-test we found no significant difference ($t(50) = -0.59$, $p = .56$, *Cohen's* $d = 0.09$) in participants' *Emotions* between the audio-visual 2D ($M = 3.27$, $SD = 0.44$) and audio-visual 3D modalities ($M = 3.31$, $SD = 0.54$). Our analysis of the sub-scales revealed that participants felt more *Anxiety* ($p = .04$, $r = -0.29$) when speaking with a 2D testimony ($M = 3.49$, $SD = 1.07$) instead of a 3D testimony ($M = 3.20$, $SD = 1.06$). Additionally, the 3D IDT ($M = 2.27$, $SD = 1.34$) inspired more *Awe* ($p = .02$, $r = 0.34$) than the 2D IDT ($M = 1.67$, $SD = 0.79$). Yet, most participants (49%) rated the 2D display mode more emotive than the 3D mode, with 29% rating 3D higher and the remaining 22% providing a balanced score. Participants who reported a strong emotional difference between both display modes in favor of the 2D display described audio-visual 3D as irritating and without any additional value in comparison to audio-visual 2D.

One participant skipped the SSQ for the 3D version. We thus only considered the remaining 50 participants for the evaluation of discomfort. Our results show significantly more *Discomfort* ($p < 0.001$, $r = 0.69$) for interactions with the 3D IDT ($M = 2.79$, $SD = 0.56$) than with the 2D IDT ($M = 2.35$, $SD = 0.41$). In particular, participants experienced significantly more severe *Eye Strain* ($p = .01$, $r = 0.68$) during the use of 3D testimonies ($M = 3.14$, $SD = 1.28$) as opposed to during the use of 2D testimonies ($M = 1.96$, $SD = 1.14$). Interacting with the 3D IDT ($M = 1.60$, $SD = 1.01$) also caused comparatively stronger, yet overall minor, feelings of *Nausea* ($p < 0.001$, $r = 0.37$) than the 2D IDT ($M = 1.24$, $SD = 0.59$), which further explains the difference in discomfort.

For general preference of level of visual modality, 49% selected audio-visual 2D and 51% selected audio-visual 3D.

Table 5: Statistically significant results of the study across the levels audio-visual 2D vs. audio-visual 3D. * indicates p < .05 and * is p < .001.**

Variable	Statistic	Effect size	Levels comparisons (Mean, SD)
Attention	z = 2.43*	r = 0.34	3D (2.27, 1.34) > 2D (1.67, 0.79)
Anxiety	z = -2.06*	r = -0.29	2D (3.49, 1.07) > 3D (3.20, 1.06)
Awe	z = 2.43*	r = 0.34	3D (2.27, 1.34) > 2D (1.67, 0.79)
Discomfort	z = 4.85***	r = 0.69	3D (2.79, 0.56) > 2D (2.35, 0.41)
Nausea	z = 2.58*	r = 0.36	3D (1.60, 1.01) > 2D (1.24, 0.59)
Eyestrain	z = 4.78***	r = 0.68	3D (3.14, 1.28) > 2D (1.96, 1.14)

4.4 Audio-visual 2D vs. Audio-visual 3D: Qualitative Results

We used the same qualitative method as for Study 1: The first author investigated the transcribed interviews and participants' explanations on the questionnaire using manual inductive coding of primarily semantic meanings and reflexive thematic analysis [19, 20]. The approach was mainly experiential and constructionist [24]. We focused on shared meaning-based patterns of experiences related to the output modality. After the first coding iteration, we constructed candidate themes and sub-themes. We revised the codes and themes in subsequent repetitions, with the final iteration taking place after eight weeks. The resulting themes and sub-themes can be seen in Figure 8. The individual participants are listed in Table 7 in Appendix B.

4.4.1 Authenticity and engagement of IDTs: The three-dimensional display made the realism of the digital witness more plausible and the conversation more believable. This helped participants to emotionally and cognitively engage with the IDT.

Spatial depth and proximity: The additional depth cues conveyed by the 3D modality aided the perceived realism of the conversation. The increased sense of space amplified feelings of co-presence and immersion. Participants contrasted the flat appearance of the 2D IDTs with the noticeable layers of depth of the 3D IDTs. As the room virtually extended behind the armchair and the witness emerged from the screen, users felt both spatially and emotionally closer: *"With the 3D representation you had the feeling he sits in the same room and I believe that you will also remember much more, so that the memory will remain longer and you felt more emotionally connected, by the fact that you had the feeling the person was closer to you" (P20).* Besides making the interaction more authentic and enjoyable, the elevated co-presence with the 3D display supported focus and engagement: *"I completely blanked out [other people], I completely blanked out the environment, and I was virtually in the room with the person... it came very close to a real setting" (P37).* However, these experiences were not universal. For a few participants, the same depth cues appeared unnatural or provided little benefit.

Vividness: The three-dimensional IDTs were perceived as more vibrant and lifelike overall. 3D added expressiveness and detail to facial and body movements, which made the digital witnesses appear more natural and alive. After interacting with Eva Umlauf's 3D IDT, P10 recalled that *"it was much more clearly distinguished and you notice when she made larger movements, for example, when she showed her tattoo, that was just much more distinct" (P10).* P25 found the same IDT response in 2D less perceivable: *"So the 2D version was fine, let's put it that way. You could do the same thing. But it wasn't as true to the original, because you just couldn't see the hands that well and especially when the lady showed her tattoo, she could have saved herself the trouble" (P25).* While most participants found the increase in vividness and authenticity appealing, users could be drawn to and get lost in the visual details. The enhanced realism can also be overwhelming. Some participants remarked that they preferred the 2D IDT because it was less real and thus less intimidating, while others found their emotional discomfort topic appropriate and desirable.

Appealing novelty: Our participants voiced a general fascination with novel technology as part of their reasoning for why they were interested and excited during interactions with IDTs, irrespective of presentation mode. The use of new or uncommon digital formats made engaging with otherwise familiar or difficult themes overall more attractive and special. The allure of novelty also led participants to differentiate between output modalities. In direct comparison, 3D was deemed more appealing, since it was perceived as less commonplace and, thus, less mundane: *"It sparks more interest because there are not as many 3D as 2D presentations" (P32).* They qualified that the appeal due to novel technology might be more prevalent among younger users, like students, and less effective among senior users. However, several participants also emphasized that the contemporary witnesses and their testimonies shall remain at the center of the interaction and cautioned not to put the *"focus on technology instead of content" (P43).* Obtrusive interfaces could add a barrier between the user and the digital witness, resulting in a shrouded and less authentic experience.

4.4.2 Physical discomfort of 3D IDTs: Our participants often expressed that the interaction with the 3D IDTs was less comfortable than the interaction with the 2D IDTs. The strain while viewing stereoscopic images and the glasses-based implementation of the 3D display inhibited their immersion and influenced their preference in output modality.

Uncomfortable glasses: Having to wear 3D glasses was mentioned as the main reason for experiencing physical discomfort. Some participants found that the additional weight affirmed the importance of the testimony and the polarized lenses discouraged them from averting their focus from the stereoscopic display of the digital witness. They also equated the act of putting on the glasses to the deliberate decision to immerse themselves in the experience

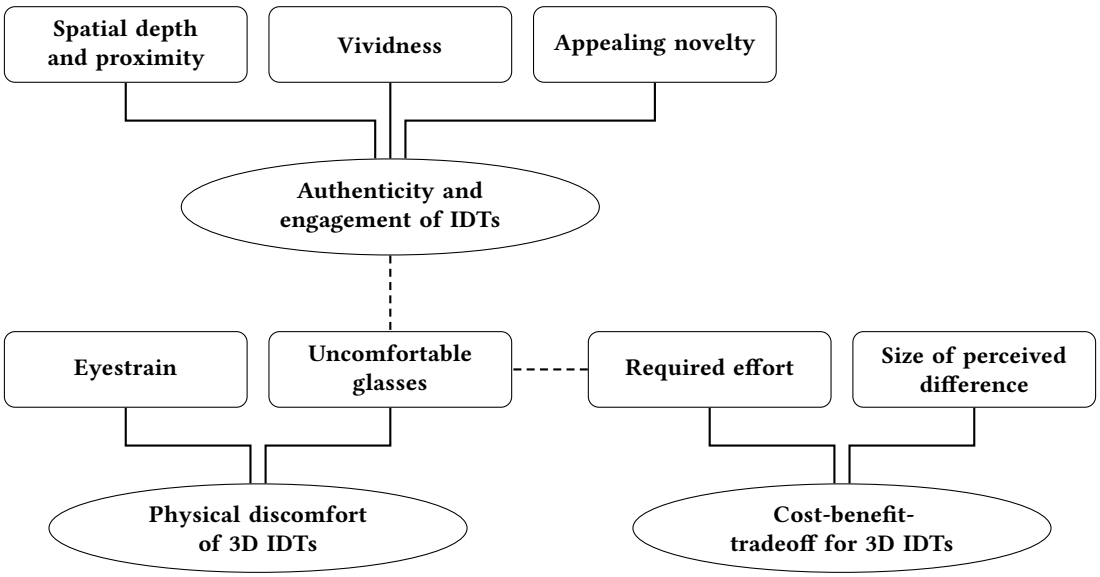


Figure 8: The themes and sub-themes derived from our analysis of interviews and free-text explanations of Study 2.

and engage with the IDT. However, several participants felt annoyed, distracted, and less present. It particularly hindered users who already had to wear vision-correcting glasses and consequently favored the 2D version: *"3D glasses are impractical for people who wear glasses and severely limit comfort. I would have preferred 3D if it weren't for the glasses problem" (P36).* The participants were confident that future advancements in display technology will solve these issues and proposed contact lenses as an interim solution.

Eyestrain: Viewing the 3D IDTs was frequently more strenuous on the eyes, which sometimes led to dizziness or headaches. Longer interaction periods were increasingly exhausting and unpleasant: *"If it had lasted a bit longer, I would have felt more comfortable in the 2D version" (P10).* To remedy these symptoms, affected participants occasionally looked away from the digital witness to rest and regain their focus, which temporarily disrupted their engagement. Consequently, employing 3D IDTs in short or custom-length sessions can mitigate physical discomfort in users.

4.4.3 Cost-benefit-tradeoff for 3D IDTs: While reflecting on their experiences, the participants also discussed the effort required for the implementation and use of the IDTs. They gathered and weighed the advantages and disadvantages of the 2D and 3D IDTs to determine their preferred modality. In many cases, these conclusions were ambiguous.

Required effort: 3D IDTs were perceived as overall more time-consuming and expensive. Our participants argued that the technical constraints and corresponding costs limit the installation and supply of IDTs at educational institutions. From a pragmatic point of view, 2D IDTs are easier to set up and could, therefore, be more accessible and available to the public. In addition to the expenses for the initial installation and technical maintenance, operating 3D IDTs requires more effort: *"I think the implementation of the 2D presentation is easier because no glasses have to be distributed. I find long fusses before a lecture annoying" (P36).* While usage of IDTs is commonly chaperoned by institutional staff already [40], 3D IDTs can add further implementation-specific steps and tasks to the interaction process.

Size of perceived difference: The participants weighed the aforementioned drawbacks and benefits, including those outlined in Section 4.4.1 and Section 4.4.2, subjectively and differently. Although some ultimately leaned strongly towards either 2D or 3D, users frequently remained undetermined with no output modality distinctly outweighing the other. Their indifference was the result of two types of user experience. The first type noticed no or only insignificant differences between 2D and 3D: *"With the first witness, I thought it was 3D anyway, and then was surprised to find it wasn't" (P12).* The second type perceived distinct advantages and disadvantages, but the aggregated difference was inconsequential: *"[T]he experience is incredibly valuable and one dimension more or less doesn't make as much of a difference" (P17).* We also found this ambivalence in preferred modality in our quantitative analysis (see Figure 7). The participants reconciled diverging views and suggested that both, 2D and 3D, should be used diligently. Dynamically adapting to user needs would improve accessibility and general user experience: *"I would like, as a suggestion for improvement, if you could let future users in schools or museums decide with the push of a button. I would leave the choice between monoscopic or stereoscopic up to the users" (P45).*

5 DISCUSSION

Our findings have several implications for the development and implementation of IDTs and lifelike ECAs. While we found no version to be unviable, we provide guidance for educational implementations at schools or museums and uncover opportunities for improvements and revisions of presently employed designs. We

discuss each study individually and conclude with design recommendations.

5.1 Audio-only vs. Audio-visual 2D × In-person vs. Online: Discussion

While a realistic human voice alone is already capable of communicating a wide range of emotions, corresponding visual representations add further social cues and visual behavioral signals. Users strongly prefer verbally addressing virtual humans face-to-face, which further supports CASA [73] and MRT [30]. While our quantitative analysis found no statistically significant difference in presence and emotional state between modalities, our participants reported that the visual display fosters focus, emotional connection, and engagement. They characterized interactions with a visually embodied IDT as more believable, as the visual representation increases authenticity and allows users to recognize the witness in other contexts, such as photos and documentaries. Additionally, our participants found the 2D IDT significantly more exciting, even if we did not find a significant difference in overall user experience. This deviation from the construct is also reflected in the comparatively low internal consistency (see Table 2).

Providing social cues like body movement, gestures, and facial expressions can lead to increased quality of learning, whereas simply showing a still image of the speaker does not suffice [65]. Users who interact with audio-only IDTs tend to choose their own embodiment of their conversational partner. This substitution can include the physical hardware used during the interaction (e.g., input or output devices), which can lead to diverse associations [94] and possibly unintended effects.

The absence of visual embodiment and corresponding non-verbal cues causes challenges to user experience and interaction. Audio-only displays lead to uncertainty regarding the current state of the conversational process, like the inability to differentiate between wordless pauses during responses and the idle state waiting for a question. Responses of IDTs show immediate visual change, e.g., gestures or adjustments in posture and facial expression. Users who are denied the visual representation experience a longer delay between question and corresponding answer. This feedback is especially important for the user experience of unaccompanied users, e.g., online learners. A visual display can also offer supplementary information, like subtitles, to improve general accessibility and usability.

Since IDTs use purpose-made recordings to simulate the digital witness, the agent is auditory and visually realistic. However, this true-to-life display leads to an increase in user expectations of the conversational ability, false affordances [39], and a Habitability Gap [68]. The design has limits in available topics and is unable to reference previous questions, answers, or encounters. System "hiccups" like non-fitting answers can not be addressed and remedied with the expected human-like conversational behavior [110]. Implementations need to consider these conversational abilities since they represent a major limitation for the realism and immersiveness of IDTs.

5.2 Audio-visual 2D vs. Audio-visual 3D: Discussion

Our quantitative analysis found no significant differences in aggregated user experience, immersion, emotiveness, and user preference. This appears to stem from the constraints of the technical implementation of stereoscopic 3D in our study. Interactions with the 3D IDTs used in our study were less physically comfortable, with prolonged usage being particularly more strenuous for the eyes. Even though the stereoscopic display used passive polarized glasses, which are more pleasant than active shutter glasses [60], they were reported as a major cause of discomfort.

In contrast, during our qualitative analysis, we found that IDTs that use audio-visual stereoscopic 3D displays can increase users' feeling of being in the same room with a real and vivid witness over audio-visual 2D implementations. However, apart from the increased awe and attention and the reduced anxiety when asking questions, this is not corroborated by our quantitative findings. Consequently, in addition to use case, target audience, and technical limitations, implementations would need to explore methods of minimizing discomfort to fully utilize these potential benefits of 3D IDTs, since the physical discomfort appears to counteract advantageous effects. Users who preferred 2D were not categorically averse to 3D, citing the fact that they do not have to wear 3D glasses as the main reason. More comfortable and less obtrusive 3D implementations could result in more distinct differences in presence and overall user experience. This holds particularly true for extended or repeated use, where the discomfort becomes less tolerable over time. Autostereoscopic displays [42], for example, would require no 3D glasses and could avoid the corresponding strain.

Overall, the added discomfort of stereoscopic 3D glasses appears to not be justifiable; simpler designs using 2D presentations could suffice and be just as effective.

5.3 Design recommendations

Effective and convincing concepts require consistency in realism beyond output modality [48, 68, 72]: Increased efforts for a more realistic audio-visual embodiment need to coincide with increased efforts for a more capable and flexible CA. For IDTs and other recording-based ECAs, this affects both the planning phase prior to recording and the subsequent training phase of the NLP. A pragmatic approach to alleviating inadequate accuracy of the NLP system or a limited pool of responses is human moderators providing example questions and further context on responses or even acting as an interface between learners and IDTs. However, this additional layer also limits and restricts the agency of users during the conversation.

Tactful and unobtrusive visual displays facilitate emotional connections through immediacy [111] and improve user experience without detracting from the digital witness and their story. Implementations need to consider the surroundings and contexts of the site of operation, as these can influence the effectiveness and limit the available modalities and their levels. For example, indoor installations at museums might encounter different spatial confines, lighting conditions, and narrative structures [92] than at outdoor cultural heritage sites. Likewise, attention to potential barriers to

use, including language and pronunciation, helps to ensure accessibility. Since users' needs are diverse and potentially incompatible with other users' preferences, IDTs require customizable and adjustable output implementations in order to improve the individual user experiences. This can include the ability to switch between 2D and 3D or providing open captions or subtitles on an additional device.

Building on the "maximization" hypothesis [77] and with future advances in display technology in mind, we recommend capturing multiple modalities in high data quality for creating IDTs of contemporary witnesses. While an audio-visual 3D testimony can be converted to audio-visual 2D, audio-only, or even text, the reverse is not possible without undermining authenticity. An extensive collection of high-fidelity recordings can serve as a basis for future revisions and re-implementations to keep IDTs up-to-date, special, and appealing. At the same time, care must be taken to ensure that the respective digital medium does not eclipse the witness and their story [91]. Further design iterations can also include exploring how to account for the loss of information when not outputting all recorded modalities, like communicating the current conversational state of the IDT non-visually.

However, any IDT using sub-optimal modalities is still vastly superior to having no IDT at all. Due to their time-sensitive nature, we recommend considerate and pragmatic approaches to creating IDTs, which maximize benefits for users while keeping the corresponding costs and efforts from becoming prohibitive.

6 LIMITATIONS AND FUTURE WORK

While our work found several differences in the effects of display modality, it has some limitations: Although our studies share numerous characteristics, their concepts and methods vary. Consequently, their respective data can not be combined or compared without restrictions. With our questionnaires, we measured multiple variables using a limited number of selected items. Our qualitative evaluations returned several effects which were not identified as statistically significant by our quantitative analysis. As participants were able to decline the interviews, our qualitative data are selection-biased and may underreport impressions that participants considered not noteworthy. However, the high interview acceptance rates in both our studies limit the impact of this selection bias. We instead suspect that the items we selected for quantitative data gathering lacked sensitivity. More focused investigations of individual variables, e.g., emotional connections with the digital witness, could reveal further differences more precisely.

Our in-person studies took place in a particular setting with dimmable lights, a human-sized display, and were devoid of external distractions. Real-world implementations at schools, museums, or homes have diverse environments, which can complicate immersion. We also can not rule out that IDTs of witnesses other than the two used in our studies could have deviating effects due to their traits, e.g., their rhetorical capabilities.

Only 4% of our study participants stated that they never encountered Holocaust topics in their daily life. Our results might be less valid for users disapproving of the IDT's content or concept. As our studies were aimed at adults and 51% of our participants were university students, our results can not readily be applied to all population groups. Follow-up surveys focusing on high school students could deliver valuable insights for the modality choice of IDTs in classroom settings or study the effects of IDT modality on knowledge gain. Collaborative classroom settings can be worthwhile, as social presence increases if a user's interaction with an ECA is preceded by other users engaging with the agent [31]. Additionally, as our evaluation focused on IDTs of Holocaust survivors, our results may not be fully generalizable to IDTs on any topic. IDTs of other, potentially more joyful topics, such as a career as a musician [96] or marine biologist [74], could benefit differently from different presentation and interaction methods.

Another limitation concerns the representation of users with disabilities since our use case was not accessible to people with hearing impairments. While visually impaired people can still benefit from the embodiment of their conversational partner, current IDTs show deficits in assistive technology. Further targeted research and design revisions considering these needs are necessary.

More than half (53%) of the users in our study comparing audio-only IDTs with audio-visual 2D IDTs participated unsupervised online. This represents a potential threat to the internal validity of our findings, as we had less control over the adherence to the study procedure and the circumstances of the interaction, including possible distractors.

We displayed the audio-visual 3D IDTs with a passive 3D display which required suitable glasses. Several study participants experienced discomfort while wearing these glasses. A survey of 3D IDTs using glasses-free autostereoscopic displays [42] could further investigate the usefulness of this modality. However, human-sized autostereoscopic displays are currently less prevalent and portable than setups using projectors and glasses. Further development of immersive display technologies is necessary to solve the issue of providing as many learners as possible with widespread access to high-quality presentations of IDTs.

Future implementations of IDTs benefit from research on improving their conversational ability, by identifying and mitigating the shortcomings of finite sets of prerecorded responses [61]. They could also utilize haptic social cues [36], olfactory displays, group IDTs with multiple witnesses, contemporary witnesses of other events, or volumetric displays [35, 38, 54, 88]. In addition to the development of corresponding theoretical concepts and frameworks, comparative empirical evaluations of each of these approaches are necessary to better understand the effects on users' perception of IDTs. Although true future-proof implementations are unlikely [1], research on existing use cases is essential to advance designs and concepts for preserving the testimonies of current and future contemporary witnesses.

7 CONCLUSION

We investigated the influence of modality on user perception of IDTs, an emerging approach to digitally preserving interactive conversations with contemporary witnesses. Since IDTs are a subcategory of ECAs, they share numerous characteristics and challenges regarding immersion, user enjoyment, and emotiveness. The exclusive use of prerecorded audio and video data, as opposed to synthetic data, prevents both auditory and visual uncanniness of the virtual human. We measured multiple variables in two distinct

user studies. The first study used a between-subjects design to compare audio-only with audio-visual 2D IDTs in in-person and online settings. We found that audio-visual 2D representations provide users with a more immersive, authentic, and pleasant experience than audio-only representations. The second study used a within-subjects design to compare audio-visual 2D with audio-visual stereoscopic 3D IDTs. Our results show that audio-visual 3D IDTs are perceived as more authentic and engaging, however, advantages over experiences with audio-visual 2D IDTs are undermined by discomfort caused by 3D glasses. Since this finding is conditional on the type of 3D display, further research with alternative display types is required. Our empirical findings confirm several benefits of embodiment for CAs. The results also extend current research on false conversational affordances of audio-visually realistic human-like ECAs. We also affirm the need for IDTs to be able to dynamically adapt their interaction conditions to the user. We recommend future-oriented approaches towards digitally preserving interactive conversations with contemporary witnesses, including capturing diverse modalities in high quality. This entails considering future types of lifelike displays and technical systems which some witnesses might not live to see.

ACKNOWLEDGMENTS

We are most grateful that Eva Umlauf and Abba Naor shared their lives with us and allowed us to preserve their stories and experiences digitally. We further thank Fabian Heindl for inspiring this series of studies, Alessa Diehl for her support with our literature review on historical empathy, Saad Elbeleidy for elevating our understanding and application of qualitative research, Maxime Pedrotti and Daniel Huber for providing the video streaming architecture, Thomas Odaker and Elisabeth Mayer for conducive discussions and proofreading, and Markus Gloe and Anja Ballis for initiating and leading the LediZ project.

REFERENCES

[1] Neta Alexander. 2021. Obsolescence, Forgotten: "Survivor Holograms", Virtual Reality, and the Future of Holocaust Commemoration. *Cinergie – Il Cinema e le altre Arti* 10, 19 (2021), 57–68. https://doi.org/10.6092/issn.2280-9481/12205

[2] Hussain M. Aljaroodi, Marc T. P. Adam, Raymond Chiong, and Timm Teubner. 2019. Avatars and Embodied Agents in Experimental Information Systems Research: A Systematic Review and Conceptual Framework. *Australasian Journal of Information Systems* 23 (2019). https://doi.org/10.3127/ajis.v23i0.1841

[3] Luis Almeida, Paulo Menezes, and Jorge Dias. 2022. Telepresence Social Robotics towards Co-Presence: A Review. *Applied Sciences* 12, 11 (2022), 5557. https://doi.org/10.3390/app12115557

[4] Elisabeth André. 2011. Design and Evaluation of Embodied Conversational Agents for Educational and Advisory Software. In *Gaming and Simulations: Concepts, Methodologies, Tools and Applications*. IGI Global, Hershey, PA, USA, 668–686. https://doi.org/10.4018/978-1-60960-195-9.ch306

[5] Gabi Arrigoni and Areti Galani. 2021. Recasting witnessing in museums: digital interactive displays for dialogic remembering. *International Journal of Heritage Studies* 27, 2 (2021), 250–264. https://doi.org/10.1080/13527258.2020.1795909

[6] Ron Artstein, Anton Leuski, Heather Maio, Tomer Mor-Barak, Carla Gordon, and David Traum. 2015. How many utterances are needed to support time-offset interaction?. In *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*. https://www.aaai.org/ocs/index.php/flairs/flairs15/paper/viewpaper/10442

[7] Ron Artstein, David Traum, Oleg Alexander, Anton Leuski, Andrew Jones, Kallirroi Georgila, Paul Debevec, William Swartout, Heather Maio, and Stephen Smith. 2014. Time-offset interaction with a holocaust survivor. In *Compilation publication of IUI'14 proceedings & IUI'14 companion*, Tsvi Kuflik (Ed.). ACM, New York, NY, 163–168. https://doi.org/10.1145/2557500.2557540

[8] Jeremy N. Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence: Teleoperators and Virtual Environments* 15, 4 (2006), 359–372. https://doi.org/10.1162/pres.15.4.359

[9] Anja Ballis. 2021. Memories and MediaMemories and Media: Pinchas Gutter's Holocaust Testimonies. In *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden*, Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann (Eds.). Eckert. Dossiers, 147–166.

[10] Anja Ballis, Michele Barricelli, and Markus Gloe. 2019. Interaktive digitale 3-D-Zeugnisse und Holocaust Education – Entwicklung, Präsentation und Erforschung. In *Holocaust Education Revisited*, Anja Ballis and Markus Gloe (Eds.). Springer Fachmedien Wiesbaden, Wiesbaden, 403–436. https://doi.org/10.1007/978-3-658-24205-3{_}22

[11] Anja Ballis and Florian Duda. 2021. Zwischen Mensch und Maschine?! – Schüler*innen befragen das interaktive 3D-Zeugnis eines Holocaust-Überlebenden. *Mitteilungen des Deutschen Germanistenverbandes* 68, 3 (2021), 284–291. https://doi.org/10.14220/mdge.2021.68.3.284

[12] Anja Ballis and Markus Gloe. 2020. Interactive 3D Testimonies of Holocaust Survivors in German language. In *Holocaust Education Revisited*, Markus Gloe and Anja Ballis (Eds.). Springer Fachmedien, Wiesbaden, 343–368. https://doi.org/10.1007/978-3-658-24207-7_21

[13] Anja Ballis and Lisa Schwendemann. 2021. 'In any case, you believe him one hundred percent, everything he says.': Trustworthiness in Holocaust survivor talks with high school students in Germany. *Holocaust Studies* (2021), 1–30. https://doi.org/10.1080/17504902.2021.1915016

[14] Hanneke Bartelds, Geerte M. Savenije, and Carla van Boxtel. 2020. Students' and teachers' beliefs about historical empathy in secondary history education. *Theory & Research in Social Education* 48, 4 (2020), 529–551. https://doi.org/10.1080/00933104.2020.1808131

[15] Nora Berner. 2022. Life Stories as Memory Carriers. In *Remembrance – Responsibility – Reconciliation*, Lothar Wigger and Marie Dirnberger (Eds.). Springer Berlin Heidelberg and Imprint J.B. Metzler, Berlin, Heidelberg, 141–155. https://doi.org/10.1007/978-3-662-64185-9{_}10

[16] Christiane Bertram, Wolfgang Wagner, and Ulrich Trautwein. 2017. Learning Historical Thinking With Oral History Interviews: A Cluster Randomized Controlled Intervention Study of Oral History Interviews in History Lessons. *American Educational Research Journal* 54, 3 (2017), 444–484. https://doi.org/10.3102/0002831217694833

[17] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 456–480. https://doi.org/10.1162/105474603322761270

[18] María José Blanca Mena, Rafael Alarcón Postigo, Jaume Arnau Gras, Roser Bono Cabré, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* 29, 4 (2017), 552–557. https://doi.org/10.7334/psicothema2016.383

[19] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[20] Virginia Braun and Victoria Clarke. 2022. Conceptual and design thinking for thematic analysis. *Qualitative Psychology* 9, 1 (2022), 3–26. https://doi.org/10.1037/qup0000196

[21] Adam Brown and Deb Waterhouse-Watson. 2014. The Future of the Past: Digital Media in Holocaust Museums. *Holocaust Studies* 20, 3 (2014), 1–32. https://doi.org/10.1080/17504902.2014.11435374

[22] Jed R. Brubaker, Meredith Ringel Morris, Dylan Thomas Doyle, Casey Fiesler, Martin Gibbs, and Joanna McGrenere. 2024. AI and the Afterlife. In *CHI'24 (ACM Digital Library)*, Florian Mueller and Penny Kyburz (Eds.). The Association for Computing Machinery, New York, New York, 1–5. https://doi.org/10.1145/3613905.3636321

[23] Christina Isabel Brüning. 2019. Holocaust Education in Multicultural Classrooms. Some Insights into an Empirical Study on the Use of Digital Survivor Testimonies. In *Holocaust Education Revisited*, Anja Ballis and Markus Gloe (Eds.). Springer Fachmedien, Wiesbaden, 391–402. https://doi.org/10.1007/978-3-658-24205-3_21

[24] David Byrne. 2022. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity* 56, 3 (2022), 1391–1412. https://doi.org/10.1007/s11135-021-01182-y

[25] Ben Caldwell, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2008. Web content accessibility guidelines (WCAG) 2.0. https://www.w3.org/TR/WCAG20/

[26] Justine Cassell. 2000. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In *Embodied conversational agents*, Justine Cassell, Tim Brickmore, Lee Campbell, Hannes Vilhjálmsson, and Hao Yan (Eds.). MIT Press, Cambridge, Mass., 1–27. https://web.media.mit.edu/~cynthiab/readings/cassell-eca-00.pdf

[27] Lucy Chabot. 1990. Nixon Library Technology Lets Visitors 'Interview' Him. *Los Angeles Times (July 21st)* (1990). https://www.latimes.com/archives/la-xpm-1990-07-21-mn-346-story.html

[28] Emna Chérif and Jean-François Lemoine. 2019. Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche et Applications en Marketing (English Edition)* 34, 1 (2019), 28–47. https://doi.org/10.1177/2051570719829432

[29] Marek Czernuszenko, Dave Pape, Daniel Sandin, Tom DeFanti, Gregory L. Dawe, and Maxine D. Brown. 1997. The ImmersaDesk and Infinity Wall projection-based virtual reality displays. *ACM SIGGRAPH Computer Graphics* 31, 2 (1997), 46–49. https://doi.org/10.1145/271283.271303

[30] Richard L. Daft and Robert H. Lengel. 1986. Organizational Information Requirements, Media Richness and Structural Design. *Management Science* 32, 5 (1986), 554–571. https://doi.org/10.1287/mnsc.32.5.554

[31] Salam Daher, Kangsoo Kim, Myungho Lee, Andrew Raij, Ryan Schubert, Jeremy Bailenson, and Greg Welch. 2016. Exploring social presence transfer in real-virtual human interaction. In *2016 IEEE Virtual Reality (VR)*, Tobias Höllerer, Victoria Interrante, Anatole Lécuyer, and Evan Suma (Eds.). IEEE, Piscataway, NJ, USA, 165–166. https://doi.org/10.1109/vr.2016.7504705

[32] Stephan Diederich, Alfred Benedikt Brendel, Stefan Morana, and Lutz Kolbe. 2022. On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems* 23, 1 (2022), 96–138. https://doi.org/10.17705/1jais.00724

[33] Aaron C. Elkins and Douglas C. Derrick. 2013. The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents. *Group Decision and Negotiation* 22, 5 (2013), 897–913. https://doi.org/10.1007/s10726-012-9339-x

[34] Jason Endacott and Sarah Brooks. 2013. An Updated Theoretical and Practical Model for Promoting Historical Empathy. *Social Studies Research and Practice* 8, 1 (2013), 41–58. https://doi.org/10.1108/SSRP-01-2013-B0003

[35] Ernst Feiler, Frank Govaere, Philipp Grieß, Simon Purk, Ralf Schäfer, and Oliver Schreer. 2020. Archiving the Memory of the Holocaust. In *Culture and Computing. HCII 2020. Lecture Notes in Computer Science*, Matthias Rauterberg (Ed.), Vol. 12215. Springer, Cham, 145–155. https://doi.org/10.1007/978-3-030-50267-6_12

[36] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies* 132 (2019), 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

[37] Paul Frosh. 2018. The mouse, the screen and the Holocaust witness: Interface aesthetics and moral response. *New Media & Society* 20, 1 (2018), 351–368. https://doi.org/10.1177/1461444816663480

[38] Cayo Gamber. 2021. Emerging Technologies and the Advent of the Holocaust "Hologram". In *Emerging Technologies and the Digital Transformation of Museums and Heritage Sites (Communications in Computer and Information Science, Vol. 1432)*, Maria Shehade and Theopisti Stylianou-Lambert (Eds.). Springer International Publishing, Cham, 217–231. https://doi.org/10.1007/978-3-030-83647-4{_}15

[39] William W. Gaver. 1991. Technology affordances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Scott P. Robertson (Ed.). ACM, New York, NY, 79–84. https://doi.org/10.1145/108844.108856

[40] Markus Gloe. 2021. Digital Interactive 2D/3D Testimonies in Holocaust Museums in the United States and Europe. In *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden*, Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann (Eds.). Eckert. Dossiers, 130–146.

[41] Erving Goffman. 1963. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. Simon & Schuster, New York.

[42] Daniel Gotsch, Xujing Zhang, Timothy Merritt, and Roel Vertegaal. 2018. TeleHuman2: A Cylindrical Light Field Teleconferencing System for Life-size 3D Human Telepresence. In *Proceedings of the 2018 SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. https://doi.org/10.1145/3173574.3174096

[43] J. Gratch, J. Rickel, E. Andre, J. Cassell, E. Petajan, and N. Badler. 2002. Creating interactive virtual humans: some assembly required. *IEEE Intelligent Systems* 17, 4 (2002), 54–63. https://doi.org/10.1109/mis.2002.1024753

[44] Joakim Gustafson, Nikolaj Lindberg, and Magnus Lundeberg. 1999. The August spoken dialogue system. In *6th European Conference on Speech Communication and Technology*. 1151–1154. https://www.isca-speech.org/archive/eurospeech_1999/gustafson99_eurospeech.html

[45] Joandi Hartendorp, Nicole Immler, and Hans Alma. 2023. Multi-perspectivity and the risk of perpetration minimisation in Dutch Holocaust and slavery education. *Journal of Curriculum Studies* 55, 6 (2023), 700–719. https://doi.org/10.1080/00220272.2023.2261998

[46] Susan Hogervorst. 2020. The era of the user. Testimonies in the digital age. *Rethinking History* 24, 2 (2020), 169–183. https://doi.org/10.1080/13642529.2020.1757333

[47] Tim Huijgen, Carla van Boxtel, Wim van de Grift, and Paul Holthuis. 2017. Toward Historical Perspective Taking: Students' Reasoning When Contextualizing the Actions of People in the Past. *Theory & Research in Social Education* 45, 1 (2017), 110–144. https://doi.org/10.1080/00933104.2016.1208597

[48] Katherine Isbister and Clifford Nass. 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies* 53, 2 (2000), 251–267. https://doi.org/10.1006/ijhc.2000.0368

[49] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies* 66, 9 (2008), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

[50] Rusty Kennedy. 2005. CMU puts words in Ben Franklin's mouth. *Pittsburgh Post-Gazette (June 30th)* (2005). https://www.post-gazette.com/news/science/2005/06/30/CMU-puts-words-in-Ben-Franklin-s-mouth/stories/200506300463

[51] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

[52] Daniel Kolb and Dieter August Kranzlmüller. 2021. Preserving Conversations with Contemporary Holocaust Witnesses: Evaluation of Interactions with a Digital 3D Testimony. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451777

[53] Katharina Kühne, Martin H. Fischer, and Yuefang Zhou. 2020. The Human Takes It All: Humanlike Synthesized Voices Are Perceived as Less Eerie and More Likable. Evidence From a Subjective Ratings Study. *Frontiers in Neurorobotics* 14 (2020). https://doi.org/10.3389/fnbot.2020.593732

[54] Yimeng Liu, Jacob Ritchie, Sven Kratz, Misha Sra, Brian A. Smith, Andrés Monroy-Hernández, and Rajan Vaish. 2023. Memento Player: Shared Multi-Perspective Playback of Volumetrically-Captured Moments in Augmented Reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. https://doi.org/10.1145/3544549.3585588

[55] Kate Loveys, Gabrielle Sebaratnam, Mark Sagar, and Elizabeth Broadbent. 2020. The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review. *International Journal of Social Robotics* 12, 6 (2020), 1293–1312. https://doi.org/10.1007/s12369-020-00680-7

[56] Minhua Ma, Sarah Coward, and Chris Walker. 2017. Question-Answering Virtual Humans Based on Pre-recorded Testimonies for Holocaust Education. In *Serious Games and Edutainment Applications*. Springer, Cham, 391–409. https://doi.org/10.1007/978-3-319-51645-5_18

[57] Karl F. MacDorman and Hiroshi Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies* 7, 3 (2006), 297–337. https://doi.org/10.1075/is.7.3.03mac

[58] Mykola Makhortykh. 2024. AI and the Holocaust: rewriting history? The impact of artificial intelligence on understanding the Holocaust. (2024). https://doi.org/10.54675/ZHJC6844

[59] Guido Makransky, Philip Wismer, and Richard E. Mayer. 2019. A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning* 35, 3 (2019), 349–358. https://doi.org/10.1111/jcal.12335

[60] Aamir Saeed Malik, Khairuddin, Raja Nur Hamizah Raja, Hafeez Ullah Amin, Mark Llewellyn Smith, Nidal Kamel, Jafri Malin Abdullah, Samar Mohammad Fawzy, and Seongo Shim. 2015. EEG based evaluation of stereoscopic 3D displays for viewer discomfort. *BioMedical Engineering OnLine* 14, 1 (2015), 21. https://doi.org/10.1186/s12938-015-0006-8

[61] Alan S. Marcus, Rotem Maor, Ian M. McGregor, Gary Mills, Simone Schweber, Jeremy Stoddard, and David Hicks. 2022. Holocaust education in transition from live to virtual survivor testimony: pedagogical and ethical dilemmas. *Holocaust Studies* 28, 3 (2022), 279–301. https://doi.org/10.1080/17504902.2021.1979176

[62] Donald Marinelli and Scott Stevens. 1998. Synthetic Interviews: The Art of Creating a 'Dyad' between Humans and Machine-Based Characters. In *Proceedings of the Sixth ACM International Conference on Multimedia: Technologies for Interactive Movies (MULTIMEDIA '98)*. Association for Computing Machinery, New York, NY, USA, 11–16. https://doi.org/10.1145/306774.306780

[63] Michael Massimi, Will Odom, David Kirk, and Richard Banks. 2010. HCI at the end of life. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA. https://doi.org/10.1145/1753846.1754178

[64] Scott E. Maxwell, Harold D. Delaney, and Ken Kelley. 2018. *Designing experiments and analyzing data: A model comparison perspective* (third edition ed.). Routledge, New York and London. https://doi.org/10.4324/9781315642956

[65] Richard E. Mayer. 2014. Principles Based on Social Cues in Multimedia Learning: Personalization, Voice, Image, and Embodiment Principles. In *The Cambridge Handbook of Multimedia Learning*, Richard E. Mayer (Ed.). Cambridge University Press, New York, NY, 345–368. https://doi.org/10.1017/CBO9781139547369.017

[66] Wade J. Mitchell, Kevin A. Szerszen, Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. MacDorman. 2011. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* 2, 1 (2011), 10–12. https://doi.org/10.1068/i0415

[67] Roger K. Moore. 2017. Appropriate voices for artefacts: Some key insights. In *Proceedings of the 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots*, Angela Dassow, Ricard Marxer, and

Roger K. Moore (Eds.). 7–11. https://vihar-2017.vihar.org/assets/vihar2017_proceedings.pdf

[68] Roger K. Moore. 2017. Is Spoken Language All-or-Nothing? Implications for Future Speech-Based Human-Machine Interaction. In *Dialogues with social robots*, Kristiina Jokinen and Graham Wilcock (Eds.). Springer, Singapore, 281–291. https://doi.org/10.1007/978-981-10-2585-3_22

[69] Masahiro Mori. 1970. Bukimi no tani genshō [the uncanny valley]. *Energy* 7, 4 (1970), 33–35.

[70] Christos N. Moridis and Anastasios A. Economides. 2012. Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions. *IEEE Transactions on Affective Computing* 3, 3 (2012), 260–272. https://doi.org/10.1109/t-affc.2012.6

[71] Muhanna A. Muhanna. 2015. Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions. *Journal of King Saud University - Computer and Information Sciences* 27, 3 (2015), 344–361. https://doi.org/10.1016/j.jksuci.2014.03.023

[72] Clifford Nass and Li Gong. 1999. Maximized modality or constrained consistency?. In *International Conference on Auditory-Visual Speech Processing*, D. W. Massaro (Ed.). https://www.isca-speech.org/archive_open/avsp99/av99_001.html

[73] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Beth Adelson (Ed.). ACM, New York, NY, 72–78. https://doi.org/10.1145/191666.191703

[74] National Marine Mammal Foundation. 2022. Dr. Sam Ridgway: Storyfile Experience. https://www.nmmf.org/sam-ridgway-storyfile-experience/ (visited on 2024-13-03).

[75] Anika Nissen, Colin Conrad, and Aaron Newman. 2023. Are You Human? Investigating the Perceptions and Evaluations of Virtual Versus Human Instagram Influencers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. https://doi.org/10.1145/3544548.3580943

[76] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (2018). https://doi.org/10.3389/frobt.2018.00114

[77] Dhaval Parmar, Stefan Olafsson, Dina Utami, Prasanth Murali, and Timothy Bickmore. 2022. Designing Empathic Virtual Agents: Manipulating Animation, Voice, Rendering, and Empathy to Create Persuasive Agents. *Autonomous Agents and Multi-Agent Systems* 36, 1 (2022), 1–24. https://doi.org/10.1007/s10458-021-09539-1

[78] Pat Pataranutaporn, Joanne Leong, Valdemar Danry, Alyssa P. Lawson, Pattie Maes, and Misha Sra. 2022. AI-Generated Virtual Instructors Based on Liked or Admired People Can Improve Motivation and Foster Positive Emotions for Learning. In *2022 IEEE Frontiers in Education Conference (FIE)*. IEEE. https://doi.org/10.1109/fie56618.2022.9962478

[79] Reinhard Pekrun, Thomas Goetz, Anne C. Frenzel, Petra Barchfeld, and Raymond P. Perry. 2011. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology* 36, 1 (2011), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

[80] Gustav Bøg Petersen, Aske Mottelson, and Guido Makransky. 2021. Pedagogical Agents in Educational VR: An in the Wild Study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. https://doi.org/10.1145/3411764.3445760

[81] Rachel Phinnemore, Mohi Reza, Blaine Lewis, Karthik Mahadevan, Bryan Wang, Michelle Annett, and Daniel Wigdor. 2023. Creepy Assistant: Development and Validation of a Scale to Measure the Perceived Creepiness of Voice Assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. https://doi.org/10.1145/3544548.3581346

[82] Delphine Potdevin, Céline Clavel, and Nicolas Sabouret. 2021. Virtual intimacy in human-embodied conversational agent interactions: the influence of multimodality on its perception. *Journal on Multimodal User Interfaces* 15, 1 (2021), 25–43. https://doi.org/10.1007/s12193-020-00337-9

[83] Sheizf Rafaeli. 1988. Interactivity: From new media to communication. *Sage annual review of communication research: Advancing communication science* 16 (1988), 110–134.

[84] Jasia Reichardt. 1978. *Robots: Fact, fiction, and prediction.* Thames and Hudson, London.

[85] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. 2021. Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *International Journal of Human–Computer Interaction* 37, 1 (2021), 81–96. https://doi.org/10.1080/10447318.2020.1807710

[86] Alasdair Richardson. 2021. Touching distance: young people's reflections on hearing testimony from Holocaust survivors. *Journal of Modern Jewish Studies* 20, 3 (2021), 315–338. https://doi.org/10.1080/14725886.2021.1874692

[87] Rufat Rzayev, Gürkan Karaman, Katrin Wolf, Niels Henze, and Valentin Schwind. 2019. The Effect of Presence and Appearance of Guides in Virtual Reality Exhibitions. In *Proceedings of Mensch und Computer 2019*. ACM, New York, NY, USA, 11–20. https://doi.org/10.1145/3340764.3340802

[88] Oliver Schreer, Markus Worchel, Rodrigo Diaz, Sylvain Renault, Wieland Morgenstern, Ingo Feldmann, Marcus Zepp, Anna Hilsmann, and Peter Eisert. 2022. Preserving Memories of Contemporary Witnesses Using Volumetric Video. *i-com* 21, 1 (2022), 71–82. https://doi.org/10.1515/icom-2022-0015

[89] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103–108. https://doi.org/10.9781/ijimai.2017.09.001

[90] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (2001), 266–281. https://doi.org/10.1162/105474601300343603

[91] Corey Kai Nelson Schultz. 2021. Creating the 'virtual' witness: the limits of empathy. *Museum Management and Curatorship* (2021), 1–16. https://doi.org/10.1080/09647775.2021.1954980

[92] Jae-Eun Shin and Woontack Woo. 2023. How Space is Told: Linking Trajectory, Narrative, and Intent in Augmented Reality Storytelling for Cultural Heritage Sites. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. https://doi.org/10.1145/3544548.3581414

[93] John Short, Ederyn Williams, and Bruce Christie. 1976. *The social psychology of telecommunications.* Wiley, London.

[94] Michael Shorter, Bettina Minder, Jon Rogers, Matthias Baldauf, Aurelio Todisco, Sabine Junginger, Aysun Aytaç, and Patricia Wolf. 2022. Materialising the Immaterial: Provotyping to Explore Voice Assistant Complexities. In *Designing Interactive Systems Conference 2022 (ACM Digital Library)*, Florian Mueller (Ed.). Association for Computing Machinery, New York,NY,United States, 1512–1524. https://doi.org/10.1145/3532106.3533519

[95] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364, 1535 (2009), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

[96] Etan Smallman. 24.3.2021. Freak out! It's Nile Rodgers in your living room, singing and answering questions. *The Guardian* (24.3.2021). https://www.theguardian.com/artanddesign/2021/mar/24/freak-out-nile-rodgers-digital-portrait-living-room-singing-answering-questions

[97] Lisa Spiro, Janice Bordeaux, Diane Butler, Andrea Martin, Chris Pound, Aniko Sandor, and Geneva Henry. 2005. The Shoah Visual History Archive: Experience from the Classroom. In *Proceedings of ED-Media 2005, World Conference on Educational Multimedia, Hypermedia & Telecommunications*, Piet Kommers (Ed.). Association for the Advancement of Computing in Education (AACE), 355–360. https://www.learntechlib.org/p/20107/

[98] Sanna Stegmaier and Svetlana Ushakova. 2021. The Production of German- and Russian-Language Interactive Biographies: (Trans)National Holocaust Memory between the Broadcast and Hyperconnective Ages. In *Digital Holocaust Memory, Education and Research*. Palgrave Macmillan, Cham, 61–96. https://doi.org/10.1007/978-3-030-83496-8{_}4

[99] S. Shyam Sundar, Haiyan Jia, T. Franklin Waddell, and Yan Huang. 2015. Toward a Theory of Interactive Media Effects (TIME): Four models for explaining how interface features affect user psychology. In *The Handbook of the Psychology of Communication Technology*, S. Shyam Sundar (Ed.). John Wiley & Sons, Ltd, Hoboken, New Jersey, 47–86. https://doi.org/10.1002/9781118426456.ch3

[100] Katy Tcha-Tokey, Olivier Christmann, Emilie Loup-Escande, and Simon Richir. 2016. Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *The International Journal of Virtual Reality* 16, 1 (2016), 33–48. https://193.48.193.34/handle/10985/11352

[101] Angela Tinwell, Mark Grimshaw, and Deborah Abdel Nabi. 2015. The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters. *International Journal of Mechanisms and Robotic Systems* 2, 2 (2015), 97–110. https://doi.org/10.1504/IJMRS.2015.068991

[102] Angela Tinwell, Mark Grimshaw, Deborah Abdel Nabi, and Andrew Williams. 2011. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior* 27, 2 (2011), 741–749. https://doi.org/10.1016/j.chb.2010.10.018

[103] David Traum, Andrew Jones, Kia Hays, Heather Maio, Oleg Alexander, Ron Artstein, Paul Debevec, Alesia Gainer, Kallirroi Georgila, Kathleen Haase, Karen Jungblut, Anton Leuski, Stephen Smith, and William Swartout. 2015. New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling. In *International Conference on Interactive Digital Storytelling (Lecture notes in computer science, Vol. 9445)*. Springer, Cham, 269–281. https://doi.org/10.1007/978-3-319-27036-4_26

[104] Italo Trizano-Hermosilla and Jesús M. Alvarado. 2016. Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology* 7 (2016), 769. https://doi.org/10.3389/fpsyg.2016.00769

[105] Brenda Trofanenko. 2017. "We Tell Stories": Oral History as a Pedagogical Encounter. In *Oral History and Education*, Kristina R. Llewellyn and Nicholas Ng-A-Fook (Eds.). Palgrave Macmillan, New York, NY, USA, 149–165. https://doi.org/10.1057/978-1-349-95019-5_8

[106] USC Shoah Foundation. 01.08.2022. Dimensions in Testimony Reaches Milestone of 50 Interactive Interviews. https://sfi.usc.edu/news/2021/12/32271-dimensions-testimony-reaches-milestone-50-interactive-interviews (visited on 2023-07-14).

[107] Janie A. van Dijk, Mirjam J. A. Schoutrop, and Philip Spinhoven. 2003. Testimony therapy: treatment method for traumatized victims of organized violence. *American Journal of Psychotherapy* 57, 3 (2003), 361–373. https://doi.org/10.1176/appi.psychotherapy.2003.57.3.361

[108] Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. 2005. Building characters: Lessons drawn from virtual environments. In *Proceedings of toward social mechanisms of android science: A CogSci 2005 workshop*. 119–126.

[109] Valentijn T. Visch, Ed S. Tan, and Dylan Molenaar. 2010. The emotional and cognitive effect of immersion in film viewing. *Cognition & Emotion* 24, 8 (2010), 1439–1445. https://doi.org/10.1080/02699930903498186

[110] Sruthi Viswanathan, Fabien Guillot, Minsuk Chang, Antonietta Maria Grasso, and Jean-Michel Renders. 2022. Addressing Hiccups in Conversations with Recommender Systems. In *Designing Interactive Systems Conference 2022 (ACM Digital Library)*, Florian Mueller (Ed.). Association for Computing Machinery, New York,NY,United States, 1243–1259. https://doi.org/10.1145/3532106.3533491

[111] Caroline Wake. 2013. Regarding the recording: the viewer of video testimony, the complexity of copresence and the possibility of tertiary witnessing. *History & Memory* 25, 1 (2013), 111–144. https://doi.org/10.2979/histmemo.25.1.111

[112] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54, 6 (1988), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

[113] Annette Wieviorka. 2006. *The era of the witness*. Cornell University Press, Ithaca and London.

[114] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225–240. https://doi.org/10.1162/105474698565686

[115] Brian Woodhouse and Paul H. Jackson. 1977. Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika* 42, 4 (1977), 579–591. https://doi.org/10.1007/BF02295980

[116] Shun-nan Yang, Tawny Schlieski, Brent Selmins, Scott C. Cooper, Rina A. Doherty, Philip J. Corriveau, and James E. Sheedy. 2012. Stereoscopic viewing and reported perceived immersion and symptoms. *Optometry and Vision Science* 89, 7 (2012), 1068–1080. https://doi.org/10.1097/OPX.0b013e31825da430

[117] Qian Yu, Tonya Nguyen, Soravis Prakkamakul, and Niloufar Salehi. 2019. "I Almost Fell in Love with a Machine": Speaking with Computers Affects Self-disclosure. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3290607.3312918

[118] Naim Zierau, Edona Elshan, Camillo Visini, and Andreas Janson. 2020. A Review of the Empirical Literature on Conversational Agents and Future Research Directions. In *International Conference on Information Systems (ICIS)*. https://www.alexandria.unisg.ch/261081/

A STUDY 1: PARTICIPANT TABLES

Table 6: Participants by participation method, as well as the combination of display modality and digital witness they interacted with.

(a) In-person participants

ID	Modality	IDT of
P1	Audio-only	Eva Umlauf
P2	Audio-only	Eva Umlauf
P3	Audio-only	Abba Naor
P4	Audio-only	Abba Naor
P5	Audio-only	Abba Naor
P6	Audio-only	Eva Umlauf
P7	Audio-only	Eva Umlauf
P8	Audio-only	Abba Naor
P9	Audio-only	Abba Naor
P10	Audio-only	Abba Naor
P11	Audio-only	Abba Naor
P12	Audio-only	Abba Naor
P13	Audio-only	Eva Umlauf
P14	Audio-only	Eva Umlauf
P15	Audio-only	Abba Naor
P16	Audio-only	Abba Naor
P17	Audio-only	Eva Umlauf
P18	Audio-visual 2D	Eva Umlauf
P19	Audio-visual 2D	Eva Umlauf
P20	Audio-visual 2D	Eva Umlauf
P21	Audio-visual 2D	Eva Umlauf
P22	Audio-visual 2D	Abba Naor
P23	Audio-visual 2D	Abba Naor
P24	Audio-visual 2D	Abba Naor
P25	Audio-visual 2D	Abba Naor
P26	Audio-visual 2D	Abba Naor
P27	Audio-visual 2D	Abba Naor
P28	Audio-visual 2D	Abba Naor
P29	Audio-visual 2D	Abba Naor
P30	Audio-visual 2D	Abba Naor
P31	Audio-visual 2D	Eva Umlauf
P32	Audio-visual 2D	Eva Umlauf
P33	Audio-visual 2D	Eva Umlauf
P34	Audio-visual 2D	Eva Umlauf
P35	Audio-visual 2D	Eva Umlauf
P36	Audio-visual 2D	Eva Umlauf
P37	Audio-visual 2D	Eva Umlauf
P38	Audio-visual 2D	Eva Umlauf
P39	Audio-visual 2D	Eva Umlauf
P40	Audio-visual 2D	Eva Umlauf

(b) Online participants

ID	Modality	IDT of
P41	Audio-only	Eva Umlauf
P42	Audio-only	Abba Naor
P43	Audio-only	Abba Naor
P44	Audio-only	Abba Naor
P45	Audio-only	Eva Umlauf
P46	Audio-only	Abba Naor
P47	Audio-only	Eva Umlauf
P48	Audio-only	Abba Naor
P49	Audio-only	Eva Umlauf
P50	Audio-only	Abba Naor
P51	Audio-only	Eva Umlauf
P52	Audio-only	Eva Umlauf
P53	Audio-only	Abba Naor
P54	Audio-only	Eva Umlauf
P55	Audio-only	Eva Umlauf
P56	Audio-only	Abba Naor
P57	Audio-only	Abba Naor
P58	Audio-only	Abba Naor
P59	Audio-only	Eva Umlauf
P60	Audio-visual 2D	Eva Umlauf
P61	Audio-visual 2D	Abba Naor
P62	Audio-visual 2D	Abba Naor
P63	Audio-visual 2D	Eva Umlauf
P64	Audio-visual 2D	Eva Umlauf
P65	Audio-visual 2D	Abba Naor
P66	Audio-visual 2D	Abba Naor
P67	Audio-visual 2D	Eva Umlauf
P68	Audio-visual 2D	Abba Naor
P69	Audio-visual 2D	Eva Umlauf
P70	Audio-visual 2D	Abba Naor
P71	Audio-visual 2D	Abba Naor
P72	Audio-visual 2D	Eva Umlauf
P73	Audio-visual 2D	Abba Naor
P74	Audio-visual 2D	Abba Naor
P75	Audio-visual 2D	Eva Umlauf
P76	Audio-visual 2D	Eva Umlauf
P77	Audio-visual 2D	Abba Naor
P78	Audio-visual 2D	Eva Umlauf
P79	Audio-visual 2D	Abba Naor
P80	Audio-visual 2D	Abba Naor
P81	Audio-visual 2D	Eva Umlauf
P82	Audio-visual 2D	Abba Naor

B STUDY 2: PARTICIPANT TABLE

Table 7: Participants by the sequence of the combinations of display modality and digital witness they interacted with.

ID	First Modality	First IDT of	Second Modality	Second IDT of
P1	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P2	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P3	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P4	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P5	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P6	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P7	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P8	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P9	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P10	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P11	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P12	Audio-visual 2D	Abba Naor	Audio-visual 3D	Eva Umlauf
P13	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P14	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P15	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P16	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P17	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P18	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P19	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P20	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P21	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P22	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P23	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P24	Audio-visual 2D	Eva Umlauf	Audio-visual 3D	Abba Naor
P25	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P26	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P27	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P28	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P29	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P30	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P31	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P32	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P33	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P34	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P35	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P36	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P37	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P38	Audio-visual 3D	Abba Naor	Audio-visual 2D	Eva Umlauf
P39	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P40	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P41	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P42	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P43	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P44	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P45	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P46	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P47	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P48	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P49	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P50	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor
P51	Audio-visual 3D	Eva Umlauf	Audio-visual 2D	Abba Naor