LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

**Dissertation**

# Design Choices for Interactive Digital Testimonies to Preserve Conversations with Contemporary Witnesses

Daniel Benjamin Kolb

2024

# Design Choices for Interactive Digital Testimonies to Preserve Conversations with Contemporary Witnesses

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

eingereicht von

Daniel Benjamin Kolb

am 05. September 2024

| | |
|---|---|
| 1. Gutachter: | Prof. Dr. Dieter A. Kranzlmüller |
| 2. Gutachter: | Prof. Alan S. Marcus, PhD |
| Tag der mündlichen Prüfung: | 04. Februar 2025 |

**Eidesstattliche Versicherung**

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt ist.

Kolb, Daniel

Name, Vorname

München, den 06. Februar 2025

Ort, Datum

Daniel Kolb

Unterschrift Doktorand:in

# Abstract

Contemporary witnesses are individuals who observed and lived through historical events. Interviews and conversations in which they share their personal experiences and individual knowledge can be particularly valuable for younger generations. Examples of this are encounters with Holocaust survivors in schools and museums. Such encounters with contemporary witnesses serve to supplement the education of attendees, to pass on oral history, and to perpetuate intangible cultural heritage.

As their availability is limited by the physical condition as well as the lifespan of the witness, there have been many approaches towards preserving such conversations for future generations. Unlike interactions with a conversational partner, linear applications of recorded media, like transcriptions and interview footage, provide limited engagement, emotion, and immersion. In contrast, interactive digital testimonies (IDTs) combine digital archives of purpose-made recordings, conversational agents, and immersive display technologies to preserve and provide conversations with contemporary witnesses in a lifelike manner. Creating IDTs is a complex, costly, and technologically demanding process involving important design decisions that may be difficult to reverse at a later point. At the same time, the effects of many design decisions and the properties of IDTs have not been investigated and are not well-understood.

In this thesis, we present a model of key components and attributes for the creation of IDTs, with testimonies of Holocaust survivors serving as a use case. To this end, we first conduct a comprehensive scientific analysis identifying the requirements of IDTs. This includes a qualitative investigation of the needs and experiences of stakeholders: educators, creators of hitherto existing IDTs, and contemporary witnesses themselves. We identified that current IDT implementations can be a more accessible alternative to conversations with human contemporary witnesses, but that the accessibility of IDTs is also limited in many ways. In addition, users encounter various difficulties when interacting with IDTs, for which assistance by human facilitators is a pragmatic but incomplete solution. We compile our findings into eight requirements and construct a conceptual model for IDTs capable of addressing the identified requirements. After describing an example implementation of the model, we present two distinct mixed-methods empirical user studies of implemented IDTs in general and on the impact of the chosen levels of the visual output in particular. We found that audio-visual 2D (i.e., monoscopic) IDTs provide users with a more immersive, authentic, and enjoyable experience than audio-only representations. Additionally, audio-visual stereoscopic 3D IDTs are perceived as even more authentic and engaging. However, advantages over experiences with audio-visual 2D IDTs are undermined by the discomfort caused by 3D glasses.

Our work serves as a guide for the creation of future IDTs, including IDTs of contemporary witnesses to other historical events or the incorporation of prospective input or output devices. Our findings also support further research on immersive displays of lifelike embodied conversational agents in various contexts, such as digital oral history, post-mortem sharing of memories, and commercial applications.

# Kurzfassung

Zeitzeug:innen sind Personen, die historische Ereignisse selber beobachtet und erlebt haben. Interviews und Gespräche, in denen sie ihre persönlichen Erfahrungen und ihr individuelles Wissen weitergeben, können für jüngere Generationen besonders wertvoll sein. Ein Beispiel dafür sind Begegnungen mit Holocaust-Überlebenden in Schulen und Museen. Diese Begegnungen mit Zeitzeug:innen dienen zur Ergänzung der Bildung, der Weitergabe von mündlichen Überlieferungen und der Bewahrung von immateriellem Kulturerbe.

Da diese Gelegenheiten sowohl durch die gesundheitliche Verfassung als auch auf die Lebenszeit der Zeitzeug:innen begrenzt ist, werden unterschiedliche Ansätze verfolgt, um solche Gespräche für zukünftige Generationen zu bewahren. Lineare Anwendungen von Aufzeichnungen, wie Transkriptionen und Interview-Mitschnitte, ermöglichen nur selten das Interesse, die Gefühle und das Eintauchen wie Unterhaltungen mit Gesprächspartnern. Interaktive digitale Zeugnisse (IDT, engl. "interactive digital testimonies") kombinieren speziell angefertigte Aufnahmen in digitalen Archiven, Gesprächsagenten und Technologien für immersive Darstellungen, um Gespräche mit Zeitzeug:innen auf lebensechte Weise zu bewahren und auch in Zukunft anzubieten. Die Erstellung von IDTs ist ein komplexer, kostspieliger und technologisch anspruchsvoller Prozess, bei dem entscheidende Abwägungen und Design-Entscheidungen getroffen werden, die mitunter später nur schwer rückgängig gemacht werden können. Gleichzeitig sind die Auswirkungen vieler Design-Entscheidungen und der sich daraus ergebenden Eigenschaften von IDTs noch nicht erforscht oder bekannt.

In dieser Arbeit wird ein Modell mit Schlüsselkomponenten und -merkmalen für die Erstellung von IDTs vorgestellt, wobei Zeugnisse von Holocaust-Überlebenden als Anwendungsfall dienen. Dafür führen wir zunächst eine umfassende wissenschaftliche Analyse durch, um die Anforderungen an IDTs zu ermitteln. Dazu gehört eine qualitative Untersuchung der Bedürfnisse und Erfahrungen der Beteiligten: Pädagog:innen, Ersteller:innen bisheriger IDTs sowie die Zeitzeug:innen selbst. Bisherige Implementierungen von IDTs stellen zwar eine leichter zugängliche Alternative zu Gesprächen mit menschlichen Zeitzeugen dar, jedoch weisen sie ihrerseits dennoch eigene Einschränkungen der Zugänglichkeit auf. Darüber hinaus stoßen die Nutzer bei der Interaktion mit IDTs auf verschiedene Schwierigkeiten, für die eine Begleitung durch menschliche Moderatoren eine pragmatische, aber unvollständige Lösung darstellt. Wir fassen unsere Erkenntnisse in acht Anforderungen zusammen und konstruieren ein konzeptionelles Modell für IDTs, welches den identifizierten Anforderungen gerecht wird. Im Anschluss an die Darlegung einer Beispielimplementierung dieses Modells erforschen wir die Implementierung anhand zweier empirischer Nutzerstudien, die sowohl qualitative als auch quantitative Methoden verwenden. Zusätzlich zu einer allgemeinen Untersuchung der Wirkung der implementierten IDTs auf Nutzer:innen, fokussieren sich die Studien auf die Auswirkungen der unterschiedlichen Formen der visuellen Ausgabe. Audiovisuelle 2D-IDTs bieten Nutzern ein intensiveres, authentischeres und angenehmeres Erlebnis als rein auditive Varianten von IDTs. Audiovisuelle 3D-IDTs werden darüber hinaus als noch authentischer und fesselnder wahrgenommen, allerdings werden die Vorteile gegenüber den Erfahrungen mit audiovisuellen 2D-IDTs durch die Unannahmlichkeiten, die die 3D-Brille verursacht, untergraben.

Diese Arbeit dient als Leitfaden für die Erstellung zukünftiger IDTs, einschließlich IDTs von Zeitzeugen anderer historischer Ereignisse sowie der Einbeziehung neuer Eingabe- oder Ausgabegeräte. Die Ergebnisse dieser Arbeit unterstützen weitere Forschungen zu immersiven Darstellungen von lebensechten verkörperten Gesprächsagenten in verschiedenen Situtationen. Diese erstrecken sich sowohl auf digitale Formen von mündlichen Überlieferungen und den Austausch von Erzählungen von Verstorbenen als auch kommerzielle Anwendungen.

# Contents

# Chapter 1

# Introduction

*"Wenn wir uns erinnern, wie rassisch, religiös und politisch Verfolgte, die vom sicheren Tod bedroht waren, oft vor geschlossenen Grenzen anderer Staaten standen, werden wir vor denen, die heute wirklich verfolgt sind und bei uns Schutz suchen, die Tür nicht verschließen."*
*[If we remember how racially, religiously and politically persecuted people who were threatened with certain death often faced closed borders of other states, we will not close the door to those who are truly persecuted and seek protection in our country today.]*

— Richard Karl Freiherr von Weizsäcker, *May 8th, 1985* [1]

## Contents



In this chapter, we provide the necessary background to motivate our research questions. Section 1.1 describes the importance of contemporary witnesses and virtual preservation of history. In Section 1.2 we present our research questions and the scientific contributions of this thesis, followed by the methodological approach in Section 1.3. Section 1.4 details the author's preliminary scientific publications and their relation to this thesis. The chapter closes with an overview of the structure of this thesis and the content of the following chapters in Section 1.5.

## 1.1 Motivation

Contemporary witnesses are people with first-hand experiences of historical events. Their recollections are detailed individual accounts of the numerous and diverse ways these historical events impacted lives. Their up-close encounters with history lend contemporary witnesses personal authenticity and their reports factual authenticity, which combine to a general "aura of authenticity" [Sab12; Sab22]. Consequently, interactions with these witnesses perform meaningful tasks for history lessons [BWT17; BS21; Tro17] and citizenship education [BSv20]. Engaging with contemporary witnesses and listening to their life stories, which might not be documented and available at other sources, can provide context, unique perspectives, and an affective component in history education. These three aspects – historical context, taking another's perspective, and affective connection – lead to a better cognitive

[1] https://www.bundespraesident.de/SharedDocs/Reden/DE/Richard-von-Weizsaecker/Reden/1985/05/19850508_Rede.html

and emotional understanding of historical figures and, consequently, historical empathy [EB13]. Personal experience reports help survivors cope with the trauma they endured [vSS03; Gam21; Ber22] and make historical events emotionally comprehensible and tangible [Ric21].

However, these learning opportunities are limited in time due to the advanced age and declining health of contemporary witnesses to numerous historical events. Unfortunately, such witnesses are eventually fading away, becoming too weak for long-distance travel or the draining and often emotionally difficult survivor talks. Consequently, a technology-based approach is needed to preserve the possibility of interacting with these contemporary witnesses for future generations.

### 1.1.1 Virtual Avatars in Popular Culture

The concept of virtual "copies" of humans as a means to "transcend" death with the help of technology has long been present in works of fiction. In the 1993 finale of season six of Star Trek: The Next Generation[2], Lt. Commander Data uses the Holodeck aboard the USS Enterprise to simulate a poker game with Isaac Newton, Albert Einstein, and Stephen Hawking. Data utilizes this opportunity to converse with the three scientists and explore how they would interact with each other. Similarly, in the 2004 movie "I, Robot"[3], Detective Spooner consults a computer-generated holographic representation of the late Dr. Lanning as part of the investigation into the circumstances of the latter's death.

A less digital and more magical take on the same idea can be found in the world of Harry Potter. Magical portraits[4] replicate the manners and knowledge of the portrayed individuals and continue to do so even after their deaths. This makes the portraits, e.g., of late Hogwarts headmasters and -mistresses, a valuable posthumous source of information and guidance in the wizarding world.

In recent years, virtual avatars can also be found outside the fictional world at real-life concerts, where virtual avatars of real artists perform on stage. One particularly notable instance was the appearance of Tupac Shakur, who died 16 years prior, at the Coachella Valley Music and Art Festival in 2012 [Mos12]. This was followed by numerous digital posthumous appearances by various artists, such as Michael Jackson at the 2014 Billboard Music Awards [Gal14] and Ronnie James Dio at the Wacken Open Air 2016 [Gro16]. Subsequently, virtual avatars of deceased musicians have been used for concert tours and residencies. Without claim to completeness, this list of artists includes Maria Callas [Hui18], Roy Orbison [Kre18], Buddy Holly [Ree19a], Frank Zappa [Vir19], Whitney Houston [Ree19b], and Elvis Presley [She24]. Occasionally, there are even performances by virtual avatars while the original artists are still alive, as in the cases of ABBA [BB21] and KISS [Amo23].

So far none of these real-world virtual avatars offer the audience opportunities to interact with the depicted individuals.

### 1.1.2 Interactive Digital Testimonies

One approach to true-to-life preservation of personal encounters is the virtually simulated dialog through Interactive Digital Testimonies (IDTs). IDTs combine a database containing exclusively pre-recorded responses of the original human witness with a conversational user interface [BBG19]. They differ from linear video recordings of interviews with contemporary witnesses, like the Shoah Visual History Archive [Spi+05], by giving users control over the selection and sequence of topics. The exclusion of synthetic replies differentiates IDTs from conventional Conversational Agents (CA) that use artificially generated answers based on archival information of historical characters [Pat+23]. When IDTs are combined with visual recordings of the witness, they fall into the category of Embodied Conversational Agents (ECAs). As such, IDTs build on scientific and technological advances in Natural Language Processing (NLP) as well as recording techniques and display technology to create interactive virtual avatars of contemporary witnesses.

By combining historical contextualization, perspective-taking, and affective connections, IDTs can evoke historical empathy in users [EB13]. This both cognitive and affective engagement with historical individuals can lead to a deeper and more nuanced understanding of history [HIA23]. Especially interactions and conversations with contemporary witnesses can promote strong emotional connections, which increase the quality and enjoyment of learning [BSv20] and prevent presentism [Hui+17]. Consequently, genuine-appearing emotive IDTs can foster history and citizenship competences by continually providing opportunities for affective engagement.

[2] *Descent*, Star Trek: The Next Generation, created by Gene Roddenberry, season 6, episode 26, Paramount Pictures, 1993.

[3] *I, Robot.* Directed by Alex Proyas, performances by Will Smith and James Cromwell, 20th Century Fox, 2004.

[4] https://www.wizardingworld.com/writing-by-jk-rowling/hogwarts-portraits

However, 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 [SU21], 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. In this thesis, we consequently explore relevant design considerations for the creation of IDTs, provide required empirical research on the use of AI in Holocaust education [Mak24], and contribute to research on digital afterlife applications [Bru+24; HN24].

## 1.2 Research Questions and Contributions

After briefly introducing the concept of IDTs in the previous section, we define the main research question (**RQ**) of this thesis as follows:

> **RQ:** How can Interactive Digital Testimonies be designed to preserve conversations with contemporary witnesses?

We address **RQ** by identifying and answering the underlying subquestions (**SQ**):

**SQ-1:** What are the requirements to digitally preserve interactive conversations with contemporary witnesses?

**SQ-2:** What are the associated practices or shortcomings of current IDT implementations?

**SQ-3:** How can IDTs be designed to fulfill the identified requirements?

**SQ-4:** How does the type of visual output modality affect how users experience the interaction with the implemented IDT?

The answers to these **SQs** form the following contributions (**C**) of this thesis:

**C-1:** A systematic discussion of literature on requirements and properties of IDTs.

**C-2:** A semi-structured qualitative study of IDT experts and stakeholders.

**C-3:** A conceptual model for IDT design that implements the identified requirements.

**C-4:** An example implementation of the conceptual IDT model as a proof of concept.

**C-5:** An empirical evaluation of the implemented IDT in general and its visual modality in particular.

## 1.3 Methodical Approach

We use both theoretical and empirical methods to answer the research questions posed above. Our approach is illustrated in Figure 1.1. It mirrors one iteration of the Human-Centered Design process (ISO 9241-210:2019) [Int19], which itself implements the Shewhart Cycle [SD39] and, by extension, the Baconian scientific method [Bac20].

1. We gather requirements and relationships between requirements by reviewing and analyzing existing literature.

2. For a more comprehensive understanding of the context of use, we complement the outcome with our own findings from interviews with experts and stakeholders of the production, implementation, and use of IDTs. This allows us to create a conceptual model of components, properties, and considerations for IDT designs aiming to fulfill these requirements.

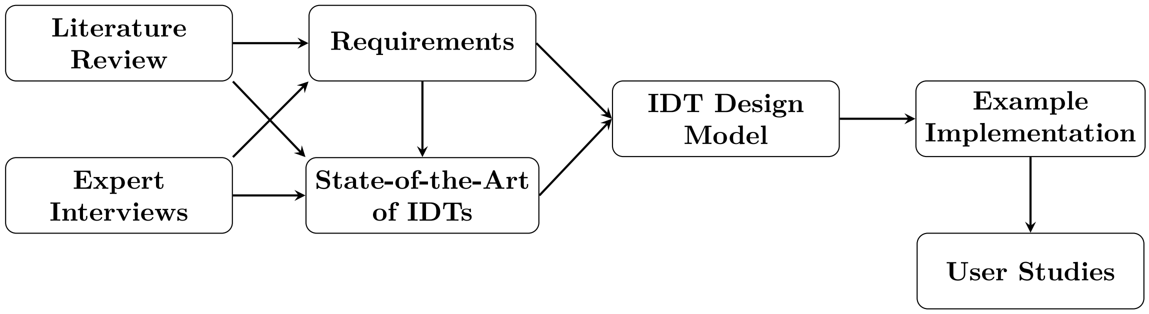



**Figure 1.1:** Methodical approach and sequence of this thesis.

3. We show how the design can be implemented and evaluate the implementation by means of user studies with a deliberate focus on the effects of its visual modality. We chose this focus since we found this aspect of IDTs to be least studied and most in need of targeted research.

Since its inception, computer science has been conceptualized as an inherently applied and interdisciplinary field of study [Fei59]. It comprises research related to the (efficient) automation of information processing [Com+89]. Since our work examines interactive computer systems for human use, it is firmly rooted in the field of computer science and draws upon its intersections with other sciences [Dix+04; BKV22; Con20], such as psychology, sociology, anthropology, ergonomics, and ethics, to investigate the topic with appropriate rigor. Consequently, we employ both quantitative and qualitative methods, as well as inductive, deductive, and abductive reasoning during our investigations.

In our research, we draw on post-positivist theory [Pop63] and acknowledge that research and researcher are not independent of each other. While we strive for objectivity, research is shaped by the theories and values of the conducting researcher [RR94] and therefore subjective. In the spirit of good scientific practice, we provide detailed and transparent descriptions of our methods and research instruments, which also encompass the author of this thesis. Consequently, this includes statements about the biases and limitations of the utilized methods as well as our own positionality [Alc88; SEG17].

During our example implementation and empirical evaluations, we focus on the use case of conversations with Holocaust survivors, as these are well-established [BF07] and widely used in numerous forms of education [GW20; Ber22]. Our investigation is situated at the tail end of the era of the witness [Wie06], when survivors of the Holocaust are still alive and able to share their life stories themselves. This group of contemporary witnesses is therefore particularly suitable for our research and deserving of the corresponding attention.

## 1.4 Author's Preliminary Scientific Work

This section contains an, at the time of writing, exhaustive list of peer-reviewed publications of the author in reverse chronological order. Each bibliographic entry is followed by a short summary of the paper, its relation to this thesis (if applicable), and a declaration of the author's contribution to the paper as defined by the 14 roles of the Contributor Roles Taxonomy CRediT[5].

In addition to these publications, the author is an active and contributing member of the scientific network "Technology Meets Testimony [TMT] – An international and interdisciplinary research network investigating the future of Holocaust survivor testimonies"[6], which is funded by the German Research Foundation DFG[7]. This network connects international scientists and experts on IDTs to facilitate the exchange of ideas and findings as well as to inspire further research projects.

### 1.4.1 Associated Publications

The following publications are connected to the research topic addressed in this thesis. Accordingly, we explain the extent to which the individual publications have contributed to this work

[5] https://credit.niso.org/
[6] https://gepris.dfg.de/gepris/projekt/448804276?language=en
[7] https://www.dfg.de/en/index.jsp

- **Daniel Kolb**. "**Design Choices for Embodied Conversational Agents to Preserve Testimonies by Contemporary Witnesses**". In: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery. Article 54, 2022. ISBN: 978-1-4503-9156-6. doi:10.1145/3491101.3503822.

**Summary:** This paper proposes research questions and corresponding methods for investigating IDTs from the perspective of HCI and the design of interactive systems. The proposal was presented, discussed, and refined at the 2022 CHI Doctoral Consortium.
**Relation to thesis:** The paper is a precursor of the research question and subquestions in Section 1.2 as well as the methodical approach in Section 1.3.
**Author's contribution:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

- **Daniel Kolb** and Dieter Kranzlmüller. "**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 (CHI EA '21). Association for Computing Machinery. Article 257, 2021. ISBN: 978-1-4503-8095-9. doi:10.1145/3411763.3451777.

**Summary:** This paper resulted from a thorough revision of the publication "**Evaluation of the Interaction with a Digital 3D Testimony – Between Emotion and Technology**", which is listed next. While the underlying user study data remained unchanged, the corresponding revised analysis, as well as the presented concepts and their implementation, provide numerous further details and insights.
**Relation to thesis:** The underlying data and the revised analysis drew our attention to the impact and relevance of the visual output modality of IDTs. It motivated our subsequent research and the design of the user studies in Chapter 6. The paper itself, however, is not part of this thesis.
**Author's contribution:** Conceptualization, Data Curation, Formal Analysis, Methodology, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

- **Daniel Kolb**. "**Evaluation of the Interaction with a Digital 3D Testimony – Between Emotion and Technology**". In: Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie. Eckert. Dossiers Vol. 1, Georg-Eckert-Institut – Leibniz-Institut für internationale Schulbuchforschung, Braunschweig, Germany, pp. 63–82, 2021. ISSN: 2191-0790. doi:10.5282/ubm/epub.75069.

**Summary:** This paper presents an exploratory user study of the user experience of the first German-speaking IDT, particularly the perceived display quality and conversational ability. It also contains a first description of the technical implementation of IDTs at the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities, as well as the corresponding user interaction design.
**Relation to thesis:** The paper and its exploratory results provided us with the first insights into user interactions with IDTs, allowing us to position and contextualize subsequent research. The paper itself, however, is not part of this thesis.
**Author's contribution:** Conceptualization, Data Curation, Formal Analysis, Methodology, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

### 1.4.2 Other Publications

The following publications are not directly associated with this thesis and are, therefore, neither used nor referenced in it. They are the result of additional research interests and efforts, primarily in the fields of Virtual Reality (VR), Augmented Reality (AR), and scientific visualization.

- Elisabeth Mayer, Rubén Jesús García Hernández, **Daniel Kolb**, Jutta Dreer, Simone Müller, Thomas Odaker, and Dieter Kranzlmüller. "**10 Years of Operating a Center with Large-Scale Virtual-Reality Installations - Developments and Learnings**". In: PRESENCE: Virtual and Augmented Reality Vol. 33. MIT Press, pp. 1-19, 2024. ISSN: 1054-7460. doi:10.1162/pres_a_00431.

**Summary:** This journal article summarizes a decade of managing a visualization center with large-scale immersive displays. We examine the advantages and obstacles of a location-based VR center, dividing the ten-year period into three stages and discussing advancements and case studies. We also share our initial experiences with a new LED CAVE and discuss future developments for VR centers.
**Author's contribution:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

- Simone Müller, **Daniel Kolb**, Matthias Müller, and Dieter Kranzlmüller. "**AI-based Density Recognition**". In: Computer Science Research Notes, Vol. 3401, No. 1. Union Agency, Science Press, 2024, pp. 227-236. ISSN: 2464-4625. doi:10.24132/CSRN.3401.24.

**Summary:** This paper proposes using artificial intelligence to analyze images and assign physical properties to identified objects. By extracting patterns from 2D images using a neural network, we can simultaneously uncover details such as volume, material type, and density. This property-based feature extraction can enhance reasoning processes, such as safe environmental motion and interaction of autonomous robots.
**Author's contribution:** Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

- Elisabeth Mayer, Thomas Odaker, **Daniel Kolb**, Simone Müller, and Dieter Kranzlmüller. "**LED CAVE - New Dimensions for Large-Scale Immersive Installations**". In: 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Institute of Electrical and Electronics Engineers, 2024, pp. 515-519. ISBN: 979-8-3503-7449-0. doi:10.1109/VRW62533.2024.00099.

**Summary:** In this paper, we present the concept of a new CAVE system based on the latest generation of LED panels. We report our preliminary findings based on 6 months of operation. This paper focuses on the implementation, the comparison to a traditional projector-based CAVE, and early user feedback. We discuss the potential future for large-scale immersive installation systems and the possibilities opened up by the advances in LED technology.
**Author's contribution:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

- Elisabeth Mayer, Rubén Jesús García Hernández, **Daniel Kolb**, Jutta Dreer, Simone Müller, Thomas Odaker, and Dieter Kranzlmüller. "**Ten years Center for Immersive Visualizations-Past, Present, and Future**". In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Institute of Electrical and Electronics Engineers, 2023, pp. 206-210. ISBN: 979-8-3503-4839-2. doi:10.1109/VRW58643.2023.00051.

**Summary:** This paper discusses the operation and learnings from maintaining a visualization center with large-scale immersive installations over the course of ten years. It includes detailed descriptions of the development processes and their changes over time as well as an outlook on the future of dedicated virtual reality centers.
**Author's contribution:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

- Fabio Genz, Niklas Fuchs, **Daniel Kolb**, Simone Müller, and Dieter Kranzlmüller. "**Evaluation of Proprietary Social VR Platforms for Use in Distance Learning**". In: Augmented Reality, Virtual Reality, and Computer Graphics. Springer International Publishing, 2021, pp. 462–480. ISBN: 978-3-030-87595-4. doi:10.1007/978-3-030-87595-4_34.

**Summary:** This paper discusses the suitability of social VR platforms for distance teaching. Combining expert interviews and an online survey, it identifies relevant criteria for teaching, which are then used to evaluate proprietary social VR platforms as well as to provide recommendations for educators.
**Author's contribution:** Conceptualization, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

- Fabio Genz, Clemens Hufeld, Simone Müller, **Daniel Kolb**, Johannes Starck, and Dieter Kranzlmüller. "**Replacing EEG Sensors by AI Based Emulation**". In: Augmented Reality, Virtual Reality, and Computer Graphics. Springer International Publishing, 2021, pp. 66–80. ISBN: 978-3-030-87595-4. doi:10.1007/978-3-030-87595-4_6.

**Summary:** This paper presents a theoretical concept and framework for reducing the number of necessary sensors measuring bio-signals. It proposes the use of pattern analysis to identify redundant sensor measurements that can instead be expressed by the combined measurements of the remaining sensors.
**Author's contribution:** Writing – Original Draft Preparation, Writing – Review & Editing

- Lea Weil, **Daniel Kolb**, Jens Weismüller, Eric Imm, and Dieter Kranzlmüller. "**Raising Awareness for Endangered Species using Augmented Reality**". In: Adjunct Proceedings of the 33rd EnviroInfo conference. Shaker Verlag GmbH, 2019, pp. 360–371. ISBN: 978-3-8440-6847-4. doi:10.5282/ubm/epub.71976.

**Summary:** This paper presents the use case of an AR mobile application to support efforts for the preservation of nature. It aims to raise awareness by displaying realistic virtual twins of otherwise reclusive and rare animals like lapwings to users. The paper also describes and discusses methods for lifelike virtual recreations of animals.
**Author's contribution:** Writing – Original Draft Preparation, Writing – Review & Editing

- **Daniel Kolb**, Wolfgang Kurtz, Jens Weismüller, Alexander von Ramm, Ralf Ludwig, and Dieter Kranzlmüller. "**Visualization of climate simulation data in virtual reality using commercial game engines**". In: Adjunct Proceedings of the 32nd EnviroInfo conference. Shaker Verlag GmbH, 2018, pp. 39–45. ISBN: 978-3-8440-6138-3. doi:10.5282/ubm/epub.71909

**Summary:** This paper describes a workflow to create a scientific VR visualization of geoscience simulation data with the help of commercial game engines. It details the challenges arising from the input data formats as well as the user interaction design.
**Author's contribution:** Conceptualization, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

## 1.5 Thesis Outline

Table 1.1 provides an outline of this thesis and an overview of which chapters answer which subquestion **SQ** with the help of which contribution **C**. Each row lists a chapter and its summary.

**Table 1.1:** Outline of the structure of the thesis and the content of each chapter as well as mapping of chapters to subquestions and contributions.

| Chapter | Content | Subquestion | Contribution |
|---|---|---|---|
| Chapter 1 | Necessary background to motivate and state the research questions as well as an overview of our methodology | | |
| Chapter 2 | Underlying concepts, state-of-the-art of IDTs, and analysis of requirements for IDTs | **SQ-1** & **SQ-2** | **C-1** |
| Chapter 3 | Qualitative study and analysis of expert interviews on the application of and interactions with IDTs | **SQ-1** & **SQ-2** | **C-2** |
| Chapter 4 | Conceptual model of IDTs and description of which properties or components are responsible for implementing which of the previously identified requirements | **SQ-3** | **C-3** |
| Chapter 5 | Presentation and discussion of how our conceptual model of IDT designs can be implemented | | **C-4** |
| Chapter 6 | Empirical evaluation of the implemented IDTs in general and their visual modality in particular through two distinct mixed-methods user studies | **SQ-4** | **C-5** |
| Chapter 7 | Outlook and implications for further research | | |
| Chapter 8 | Summary of the previous chapters | | |

# Chapter 2

# Related Work

*"If one wants to make a machine mimic the behaviour of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table. Constructing instruction tables is usually described as 'programming'."*

— Alan Turing, *October 1st, 1950* [8]

## Contents



In this chapter, we provide an overview of related work and research. In Section 2.1, we introduce underlying concepts of IDTs. We compile, analyze, and discuss the state-of-the-art of IDTs in Section 2.2. Based on our literature analysis, we formulate requirements for IDTs in Section 2.3. We examine previous models for the design of IDTs in Section 2.4. In Section 2.5, we provide a summary of this chapter.

## 2.1 Background

In this section, we introduce and explain the relevant concepts that serve as the foundations of IDTs. We discuss presence and immersion, which determine how readily users accept and engage with virtual actors or environments. Building on this, we provide an overview of considerations for the design of Embodied Conversational Agents and review approaches towards the interactive digital preservation of memories.

[8] [Tur50], p. 438

### 2.1.1 Presence and Immersion

Throughout this thesis, we use Slater's definition of presence as the feeling of "being there" in a virtual environment [Sla09] and immersion as a system property. Co-presence describes the sense of "being together" due to sensory awareness of another co-located actor [Gof63; BHB03]. Social presence, which builds on co-presence, is the perceived ability of a system to evoke intimacy and immediacy during an interaction [Gof63]. This affects the ability and willingness of users to connect and engage with the virtual actor [BHB03]. The Theory of Interactive Media Effects [Sun+15] argues that interactivity, usability, modality, realism, and social presence impact engagement and appreciation of an interactive system. The chosen display modality, i.e., how the system outputs information, 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 [VTM10] or induce detrimental side-effects like physical discomfort [Yan+12].

### 2.1.2 Embodied Conversational Agents

Conversational Agents (CAs) are computer systems designed to communicate with human beings via natural language. They build on the Computers-Are-Social-Actors (CASA) paradigm [NST94]: 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. When combined with an embodiment, such as a visual representation, CAs become Embodied Conversational Agents (ECAs). ECAs are powerful interfaces, as users typically do not have to learn new communication techniques for the face-to-face conversation. Building on simple interactions, they are also well-suited to navigate survivor testimonies as well as to strengthen user involvement [Fro18].

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 [Gra+02]. Human realism of embodied agents affects perceived presence and involvement, which influences enjoyment and trustworthiness of the conversation [Alj+19; Rhe+21]. 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 [VSS05] or creepy [Phi+23] behaviors can have detrimental effects. ECAs with not quite realistic appearances fall victim to the Uncanny Valley Effect [Mor70; MI06; Tin+11; Rei78], i.e., feelings of revulsion and aversion as well as distrust [NCN23] in users who encounter anthropomorphic yet not quite human-like depictions. The same effect occurs when body movement and speech are not aligned [TGN15] or the voice does not fit to the visual representation of the ECA [Mit+11].

The way an ECA sounds and speaks also contains social cues that benefit CASA [Fei+19]. While the effects of the visual appearance of ECAs are well substantiated, research on human-like voices is inconclusive [Zie+20]. Previous studies have shown that vocal pitch [ED13], emotional tone [ME12], and gender of voice [Yu+19] influence the way users perceive and interact with CAs. Similar to the Uncanny Valley Effect, user acceptance decreases if the realism of the voice output does not align with the system capabilities [Moo17b]. Voice design of CAs should match their affordances to raise appropriate expectations in users [Moo17a]. However, recent findings show that humans prefer realistic voices over synthesized robotic ones [KFZ20]. The perceived eeriness decreases with higher degrees of human-likeness, which challenges the existence of an auditory Uncanny Valley Effect. Further studies show that, compared to synthetic voices, human voices cause stronger feelings of trust, social presence [CL19], and intimacy [PCS21]. While aligning the realism of audio and video of human-like ECAs is beneficial [NG99; IN00], using the respectively highest quality in each modality can result in the highest overall acceptance and trustworthiness [Par+22].

Educational ECAs featuring human voice and human-like behavior cause deeper learning [May14]. 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 [Rza+19]. 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 [Bai+06]. Consequently, under some circumstances, non-embodied CAs can rival or even surpass ECAs in terms of user experience and co-presence. Although users expect human-like behavior of educational ECAs [And11], realistic behavior can have varying effects on learning and knowledge gain, depending on the type of material [PMM21]. In the case of testimonies by contemporary witnesses, we expect this impact to be comparatively minor, as these testimonies combine factual and conceptual information. Additionally, the admiration of the person portrayed by the ECA can improve interest and evoke more positive emotions [Pat+22].

We encountered increased interest in research on the user perception of visually rich ECAs, which is also reflected in a meta-analysis on the user perception of visually rich ECAs [Die+22]. The results of previous studies on the effects of realistic visual embodiment on perceived social presence [OBW18], co-presence [AMD22], and emotional connection [Lov+20] during virtual interactions did not cover ECAs using non-synthetic photorealistic videos as their representation.

### 2.1.3 Interactive Digital Preservations of Memories

Recent work investigated how testimonies and past experiences can be preserved digitally so that they can be replicated, interactively explored, and shared.

Using immersive technologies, such as VR [Sch+22] or AR [Liu+23], to display volumetric captures of events, like notable encounters with people, can allow users to repeatedly, collectively, and immersively (re-)visit the event, even if they were not present originally. While experiencing the recording, users can freely explore the scene and focus their attention on details at will. Using recordings instead of computer-generated content prevents threats to the accuracy of the reproduced event and the fidelity and naturalness of the displayed humans. However, the intricate volumetric capture process and the corresponding large size of the data limit the duration of each individual event as well as the overall duration of the recorded material. Consequently, users of current implementations rate the content provided as too short [Sch+22]. In addition, as mere observers, users of these applications cannot interact with the people depicted, which inhibits establishing an emotional connection.

By contrast, CAs using generative Artificial Intelligence (AI) can offer users such interactions with historical persons [Pat+23], which can result in increased interest and curiosity about the portrayed person. Conversely, prior familiarity with the life and background of their conversational partner can help users to ask deeper and more meaningful questions. Additionally, CAs simulating real persons are not limited to famous historical individuals and can be extended to ordinary people due to the often ready availability of extensive and suitable training data, such as professional and personal correspondence, published works, or notes. However, the application of generative AI leads to difficulties in terms of the authenticity and integrity of the presented output and the provision of a realistic audio-visual embodiment.

Memories and experiences of the past can also relate to death and grieving. Previous HCI research underlines the demand for thanatosensitive (i.e., actively engaging with death and mortality) studies on how technology and interactive systems can support people in sharing or remembering the deceased and how the bereaved appropriate existing technology to their needs [Mas+10]. This entails exploring how and what content people choose to shape their own digital heritage, as well as the attachments users form to bequeathed digital data of and by other people [KB08]. Prior investigations found that the design of systems for the bereaved should support storytelling and not view death as the end of the user's relationship with the deceased [MB11]. Additionally, (perceived) durability and personal history through repeated engagement over time contribute to the user's appreciation of such interactive systems [Odo+09].

## 2.2 State-of-the-Art of IDTs

IDTs are CAs using exclusively prerecorded responses. IDTs simulate conversations with contemporary witnesses by combining a database of pre-recorded responses, as given by the original human witness, with a conversational user interface. If the recordings include an embodiment of the contemporary witness, the corresponding IDTs are a subcategory of ECAs [MCW17], with the restriction that neither the embodiment of the agents nor their replies are artificially synthesized. This aims to create immersive and persuasive audio-visual ECAs, without the need for highly realistic computer-generated images, voices, or social cues. IDTs thereby avoid the risks of the Uncanny Valley Effect as well as hallucinating and inventing false statements about the Holocaust [Mak24]. However, the exclusion of synthetic replies and the finite number of pre-recorded responses limit conversational content and adaptability. Our use of the term IDT includes the numerous terminologies in research and media describing the same concept, such as time-offset interaction [Art+14], virtual interactive survivor testimony [Mar+22], conversational video [Ron22], conversational media [The23], or hologram [McM16].

As they provide users with a Virtual Learning Environment (VLE), IDTs can be effective tools for education. Prior research identified interactivity and immersion, which enhance motivation and engagement, as decisive aspects for the usefulness of VLEs [Mar+17; Pat+23]. Covering a variety of subjects [Kam+19], VLEs enable learners to have experiences that would otherwise be difficult or impossible in the real world. The immersive, interactive, and affective engagement with a well-produced digital copy allows for an accessible and unhindered exploration of the real original [LL11; KY18]. While encounters with a copy and an original are not equivalent, both experiences can be "auratic" and authentic. Interactivity allows users to be active learners and gain knowledge on their own terms. Immersive systems, like stereoscopic displays, can deliver a greater feeling of enjoyment while interacting and learning [Lou+17; RP19]. While non-interactive digital testimonies are of interest to students, they offer limited immersion due to the lack of interactivity [Brü19]. VLEs can gain further benefit from employing multimodality [Pap+19] and intriguing technology [Kam18].


**Figure 2.1:** Example of an IDT predecessor: Spoken dialogue system of Swedish author August Strindberg as shown in [GLL99].

Early 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 [Cha90]. 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 [GLL99] (see also Figure 2.1). Users were able to verbally ask the system about the life of the author, among other things. 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 [MS98]. 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 [Ken05], in which an image is projected in such a way that it appears to stand in front of the users. 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.

As indicated by the number of corresponding implementations, the most prominent current use case for IDTs is in Holocaust education as a means of teaching history and promoting human rights. This is also aided by the fact that considerate implementations of interactive digital media can strengthen users' engagement with survivor stories [BW14]. During our research, we encountered three different projects working on IDTs of Holocaust survivors:

- The USC[9] Shoah Foundation developed the first IDT of a Holocaust survivor in 2014 as part of its Dimensions in Testimony project [Tra+15b]. At the time of writing, they have created at least 50 IDTs, covering nine languages [USC22].

[9] University of Southern California

- The Forever Project undertakes a similar approach, producing IDTs for the National Holocaust Centre and Museum in England [MCW17].
- The German project “LediZ” (“Lernen mit digitalen Zeugnissen”, transl. “Learning with digital testimonies”) developed two German-speaking IDTs of Holocaust survivors for use in schools and museums [BG20].

The IDTs of these three projects share a number of design properties: Each testimony features a witness themself talking about their own life story. All response videos were recorded in stereoscopic three-dimensional (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 [Tra+15b; MCW17; BBG19], including an introductory witness account, an idle loop between questions, and neutral fallback responses if no matching response can be displayed.

The current design of IDTs is quasi-interactive [Raf88], as using only pre-recorded replies makes it impossible for the CA to reference user-defined verbal input or prior interactions with the IDT. While the IDTs were recorded in stereoscopic 3D, the choice of visual display, e.g., monoscopic two-dimensional (2D), Pepper’s ghost, 3D, varies by implementation and location.

In the following subsections, we provide a comprehensive descriptive overview of previous research on IDTs. We identified and subsequently dissected and analyzed a total of 29 publications that examined and investigated IDTs. In our analysis, we determined eight overarching aspects in prior research:

- Natural Speech Input
- Database of Pre-recorded Answers
- Processing of User Input
- Displaying IDT Output
- Conversational Ability
- Future-proof
- Accessibility
- Authenticity and Integrity

### 2.2.1 Natural Speech Input

IDTs utilizing voice user interfaces allow and expect user input in spoken form [Art+14; Tra16]. Freedom of choice in the structure and content of the speech input is necessary for conversations intended to be as natural and dialogical as possible [BBG19]. This flexibility also puts users at the center and in control of the interaction. It provides users with a personalized experience tailored to their respective subjective interests [Gam21; For22] and the agency to individually explore the story of the witness [BDG21; Dud21]. The natural speech input addressing IDTs can be arbitrary in meaning, syntax, or verbiage. Similar arbitrariness needs to be expected from the unrestricted sequence of user prompts, like asking questions about different topics that are unrelated to their previous query.

Current implementations usually task the users themselves with signaling the beginning and the end of their intended voice input, e.g., via push-to-talk buttons [Tra+15b]. These additional steps to the interaction can inhibit prospective users, introduce user errors [BDG21], or limit the perceived naturalness of the simulated conversation by raising the user’s awareness of the interface [SU21].

The user input can be directed or influenced by providing pre-set example questions [Mar+22] or human intermediaries (“facilitators”) who collect, adapt, and ask questions instead of the users themselves [Bal22; Glo21; AG19; AG21]. Both approaches are incompatible with the objective of independent, direct, and unmediated user interactions [Hei21; Art+15] and inhibit the user’s personal connection with the digital witness [Glo21]. We conclude that users should be able to interact with IDTs with as little external help as possible. For this reason, our analysis does not focus further on the use of human intermediaries.

### 2.2.2 Database of Pre-recorded Answers

Unlike most ECAs, the answers of IDTs are neither procedurally generated nor manually synthesized by a third party. They instead exclusively rely on databases containing purpose-made live recordings of spontaneous reactions and responses given by witnesses during structured interviews [Tra16], which also avoids potential issues related to the Uncanny Valley Effect. While the production and composition of the pre-recorded answers is not the focus of this thesis, we describe and discuss influencing factors to provide a comprehensive overview.

The pool of available answers is determined by the set of answers captured during the recording session. The interview questions, which elicit the answers from the human witness, are sourced during the preparatory pre-production stage, e.g., by identifying the most common interests and concerns of prospective users [Art+14; Art+15]. The questions are usually gathered for two categories [MCW17; Tra+15b]:

- Questions about their experiences over the course of their lives
- Personal questions like opinions and biographical information

An extensive sourcing process aims to cover as large a percentage of user input as possible with the limited amount of answers. However, unanticipated responses provided by the witness during the recording can provoke follow-up questions. Accounting for these questions requires dynamically appending the predetermined question catalog during the recording process [SH21]. The question sourcing process also needs to consider that users may dare to ask the unfeeling IDT more sensitive and distressing questions than the human witness, whose feelings they would risk to hurt [For22]. Examples for the sourcing process can be found in [Tra+15b; MCW15; MCW17; BG20; SU21].

A study of the pool of recorded replies for the IDT of Pinchas Gutter identified that 1711 responses can address 95% of unconstrained questions by users [Art+15]. However, the generalizability and transferability of this finding are limited by several factors:

- The assessment of whether a response addresses a question was made by internal annotators instead of the users themselves, who might offer deviating assessments.
- Since no automated matching took place, the 95% does not factor in operating errors or technical issues like flaws in speech recognition or inaccuracies of the matching system.
- The study also lacks details about potentially confounding factors, like the length of the interactions, the number of questions per participant, the participants' levels of knowledge, or the influence of the operation sites on the choice of questions.

Due to being a prototype, the answer pool of Pinchas Gutter's IDT was recorded over two separate recording sessions. In the same study, the responses from the first recording were found to be able to address only 58%–69% of the questions. The second recording took place five months after the first recording with the declared objective of narrowing the gap in answer coverage. Due to the accompanying increase in costs and effort, subsequent IDT production processes have not utilized multiple disconnected recording sessions with intermediate empirical analyses of gaps in answer coverage.

In order to provide an adequate answer to every possible question, an infinite set of pre-recorded answers would be necessary [Tra16]. It is still essential to supply users with a satisfactory experience with a finite set of pre-recorded answers, including a serviceable response for any given input. This warrants the production and provision of recordings with a functional purpose.

In our literature analysis, we identified three types of functional recordings for IDTs: fallback reply, idle loop, and introductory testimony.

- Fallback replies are displayed to handle input for which the database otherwise contains no suitable answer. They inform the user about the IDT's inability to provide a more suitable reply for the detected input, e.g., by acknowledging the technological limits and non-human nature of the system [Ale21]. Fallback replies can also direct users toward other resources [BBG19] or provide advice on resolving technical issues [Gam21].
- Idle loops are used to represent the idle state of the IDT, such as during the absence of input. This state occurs before a question is asked or subsequent to displaying a reply. A recording of the witness waiting and listening [Art+14] can be used to signal the idle state to the user. Since there is no theoretical upper bound for the time interval between subsequent user inputs,

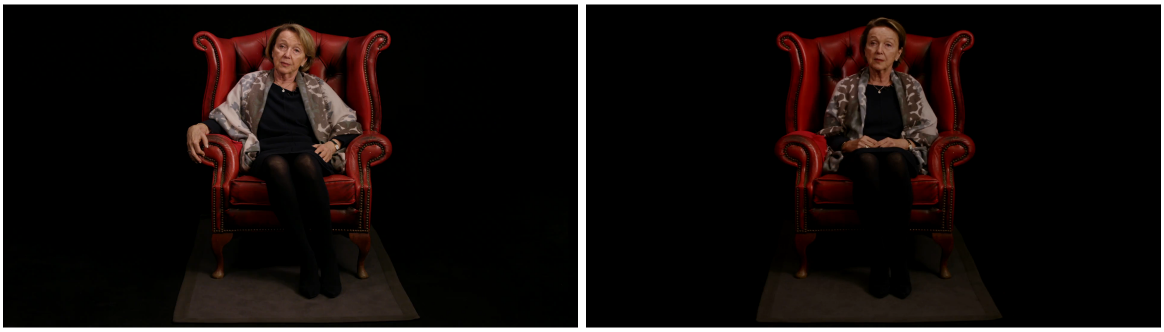


**Figure 2.2:** Example of video continuity errors within the recorded answers for a single IDT. Both images are extracted from the beginning of different answer videos. The video continuity errors include posture, hairstyle, video brightness, and lightning. Image source: LediZ project[10].

there is likewise no upper limit for the duration of displaying the recording of the witness waiting. Instead of creating and displaying excessively long idle files, short recordings are looped repeatedly. Deliberate employment of social cues in the idle loop, like an inviting stance, can further encourage users to interact with the IDT [Gam21].

- Introductory testimonies are continuous and more extensive recordings of the witness providing their testimony and personal history. IDTs mirror the structure of living testimonies, which consist of both an introductory monologue by the contemporary witness and their dialogical interaction with the audience [Bal19; BD21]. Introductory testimonies provide users with contextual information and numerous details which are instrumental in devising and formulating questions [BBG19; Gam21; SU21; Hei21]. Despite serving a specific pre-determined purpose, introductory testimonies are not exclusively functional recordings, as they allow witnesses to curate and share their experiences. Introductory testimonies can additionally be used as replies of IDTs to user prompts about the general life story of the respective contemporary witness [Bal22].

Recording multiple replies to the same prompts affords more varied and adaptive answers by the IDT to user questions. These variations can offer different levels of detail [Tra+15b] or provide diversified fallback replies [Art+15]. Recording variations of the introductory testimony preserves the ability and agency of the original human witness to dynamically adjust the content to the audience, e.g., when addressing school children instead of adults [Mar+22]. In practice, recording multiple versions requires a trade-off between producing more variations for already recorded content and expanding the limited pool of answers to cover more topics.

Since there are no constraints to the order of individual user inputs, i.e., arbitrarily switching topics, each IDT output needs to fit to any preceding output. The sequence of triggered responses frequently deviates from the order in which the witness' answers and reactions were recorded. Continuity errors, both semantical [BBG19] or intramodal [Tra+15b], between recordings would jeopardize the perceived cohesion and realism of interactions with the IDT. Semantical continuity errors can arise from references to previously recorded answers. Users should be able to understand and comprehend the output provided by the IDT without having to have experienced another specific answer beforehand. Intramodal continuity errors refer to inconsistencies in the way the answers are displayed. For example, inadvertent changes to sound volume or image brightness undermine audio continuity or video continuity, respectively. Considerations about video continuity also concern cinematic continuity, including overall consistency in camera position and perspective [BBG19] as well as the clothes and accessories worn by the witness [Tra+15b; MCW15; BBG19]. During the recording, the witnesses are instructed to start and end each answer with the same posture and facial expression [Tra+15b; MCW15; MCW17; Bal21a] to facilitate smooth transitions when displaying the answers to users. To reduce the number of outside interferences, the recordings take place in a controlled, neutral environment instead of at the homes of the witnesses. The composition of a black background and a pronounced chair equips current IDTs with a distinct visual aesthetic [Glo21; BDG21]. An example of this aesthetic as well as a selection of video continuity errors can be seen in Figure 2.2.

The limitation to pre-recorded answers constrains the ability to update and adapt the content of the IDT to temporal changes, including events experienced by the real human witness since their

[10] http://www.edu.lediz.lmu.de/

recording [Kan14]. The replies can become outdated or unable to address topics that are contemporary at the time of interaction. Remedying content gaps that appear after the recording process requires additional, incremental recording sessions. However, discontinuous recording sessions further increase the risk of continuity errors.

The process of recording high-quality answers is influenced or restricted by several environmental conditions. The welfare of the human witness during the recording sessions impacts their willingness and manner to divulge difficult or traumatic experiences [BDG21; Gam21; SH21; Mar+22]. The desired well-being is impeded by the recording taking place over multiple consecutive [Art+14] "grueling" [Gam21] days combined with the continuity requirements of wearing identical clothes and returning to the same posture and facial expression at the end of each answer. While the witnesses are not recorded in the comfort of their own homes, some even have to travel to another, potentially remote location [SH21]. This also raises ethical concerns, as IDTs exclude witnesses who are unable to endure the recording process [Gam21] or travel [Mar+22]. A close and trusting relationship with an empathic interviewer helps putting the witness at ease [SH21; SU21]. However, familiarity between the interviewer and the witness can also lead to responses presupposing knowledge that the end users might lack [Art+15]. Further limitations arise from the choice of modalities to be recorded, like stereoscopic 3D [SH21] or visual 360° [Bal19; SU21], which influence the complexity, technical requirements, and costs of the recording process [Art+15; Tra+15b; Gam21; SH21].

Prior research on IDTs does not examine or compare methods of providing the pre-recorded answers to the site of operation. Divergent technical limitations at different sites can necessitate alternative methods of hosting the database of recordings. Further research is also required on the potential need, design, and use of other functional recordings as well as non-verbal responses.

### 2.2.3 Processing of User Input

The processing of the verbal user input consists of two steps, which both utilize AI techniques [BG20; BD21]: Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU). ASR serves as a pre-processing step to increase the computer-processability, indexability, and searchability of the input data by converting the spoken words to text [BBG19; SU21]. Using this text-form, the NLU component then automatically classifies the intent of the user's input and selects a matching pre-recorded response by means of the unique identifiers of the recordings [Tra+15b; SH21]. A sophisticated automatic processing system is essential for the successful simulation of a dialogic interaction [Brü18]. Conversely, limitations to the ability to process input can inhibit users' ability to freely choose and phrase their input, diminishing their experience [BD21]. Due to the complexity of building a viable NLP system from the ground up, current implementations of IDTs instead incorporate pre-existing, proprietary software or services for ASR [Tra+15a; Dud21; SU21] or NLU [Dud21; BD21]. However, the use of closed-source software limits the traceability, transparency, and explainability [CB23] of the processing steps undertaken by the respective system components [BG20; SU21].

The training of the individual NLUs of each IDT is carried out in three phases [BG20; SU21]:

- Initialization
- Non-public test interactions
- Public use

The first phase consists of initializing the training data set with the original interview questions as well as systematic, semantic, or syntactic variations thereof [BBG19; MCW15; MCW17; SU21].

This enables first, non-public test interactions with the IDT in the second phase, which aim to train and refine the matching system [Tra+15b] and improve its initially low accuracy [Dud21]. The training process implements supervised learning [Tra16], with each question-answer-match being manually reviewed and validated or, in the case of deficient classifications, rectified [BG20; Dud21; SU21]. 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 [Dud21; SU21]. However, detailed requirements or guidelines for this decision-making process are not within the scope of this thesis.

The third phase of training the NLU is a lasting continuation of the second phase without restrictions of users. After an IDT is installed and put to use at its intended site of operation, the interactions of the end users continue to be gathered. Sustained reviews of these real-world data serve to further train and improve the accuracy of the IDT's matching system [BDG21; SU21; Gam21]. If an IDT is employed at multiple sites concurrently, a centralized NLU allows consistent and simultaneous training and adjustments for all sites [BBG19].

The NLU uses the knowledge gained during its training to build a statistical model which can assign each answer in the database a confidence value representing its probability to fit to a given input [Tra+15b; MCW17]. A confidence threshold acts as a filter since only answers with confidence values exceeding the threshold are considered suitable replies. The value of the confidence threshold impacts the IDT's ability to return suitable answers even if the user's exact wording is not present in the training data [BDG21]. If no answer surpasses the confidence threshold for a given user input, a fallback reply is returned instead. If multiple answers surpass the confidence threshold, the answer with the highest confidence is chosen [Tra+15b]. If the NLU tracks recent answers, a lower-ranked answer can be chosen instead to avoid repeating replies during a set period of time.

The extent of the delay between user input and IDT output, due to transmitting or processing data, impacts the users' impression and experience of the interaction as a fluent conversation [BBG19]. A realistic simulation of a dialogic encounter requires processing and matching user questions in real-time [MCW17; Ale21; AG21]. However, the authors of the paper refrained from further defining or specifying what magnitudes of delay constitute 'real-time' in this context. We infer that shorter processing times lead to better and more immersive user experiences. Conversely, unreliable technology and hang-ups limit the responsiveness of IDTs [Mar+22].

A key quality of the processing system is the accuracy of its question-answer-matches, as this influences whether users experience encouragement or frustration during the interaction [BG20]. Inaccurate matches may provide users with a sequence of random statements instead of constructing a cohesive meaningful narrative in the users' minds [Hei21]. The matching accuracy is not only challenged by potential NLU issues, like lack of training data [Tra+15b] or complex sentence structures [BDG21], but also flaws in the preceding processing steps. Patchy recording of user input [SU21], poor speech recognition [Tra+15b], or faulty transmission of data [BDG21] can provide the NLU component with already compromised input. The variety in user's fluency [Ale21], voice volume, accent, and dialect [Mar+22] also adds further difficulties for the ASR.

We encountered only one empirical evaluation of the matching accuracy of an IDT in operation, which found that 81.6% of answers returned by this IDT were relevant to the user's question [MCW17]. However, the examined dataset, which contained only 42 question-answer-pairs, was both small and limited. The evaluation also does not address whether the identified accuracy value is sufficient to convincingly simulate a conversation.

### 2.2.4 Displaying IDT Output

In order to emulate real encounters with witnesses [BG20], including the feeling of being in the same room [Gam21], IDTs need to reproduce the pre-recorded answers in a plausibly realistic and immersive way [Tra16; Glo21; Mar+22]. Since witnesses communicate their experiences verbally and non-verbally [Bal21a], the choice of output modalities also influences the IDT's ability to display the emotions of the witness and thereby the affective quality of the interaction [Gam21]. For this reason, current IDTs use both auditory and visual output [Tra+15b; Tra16; Bal21b; Mar+22]. Stereophonic audio output immersively delivers the prosodic features of the answers [Bal21a], like intonation, pauses, accent, or stress [MCW17]. The visual embodiment of the digital witness supplements the answers with silent cues, including posture, gestures, facial expressions, and eye contact [Art+14; MCW15; Gam21; Bal21a]. Displaying these recorded non-verbal cues provides further details about the emotional state of the witness and helps users make a personal connection [Tra+15b]. Conversely, substituting or supplementing sections of the IDT with a synthetic, computer-generated version of the witness [MCW17] compromises its overall realism, plausibility, and stimulus for empathy [Sch21].

Current IDT implementations employ and adapt to a variety of visual display methods and fidelity levels. Automultiscopic volumetric displays, which present 3D content that is observable from multiple viewpoints and without the need for auxiliary tools like glasses, would facilitate a highly realistic and natural visual presentation of the digital witness [Tra+15b]. However, appropriate displays do not exist yet [SU21]. Current IDT implementations thus frequently resort to stereoscopic displays to provide users with realistic, 3D visual presentations of witnesses [Brü18; BBG19; AG21; BD21; Bal21a]. These require users to wear 3D glasses and view the digital witness from a narrow number of angles to fully experience the 3D effect. Further utilized display methods include Pepper's Ghost [BBG19; SU21], which presents the witnesses in 2D with an additional illusion of depth, and plain 2D for, e.g., temporary installations [BBG19; Glo21], as it is less complex and less laborious to set up. For each display method, the employed visual fidelity of the video, like the resolution [Art+14; Tra+15b], can also vary with the specifications of the actual output device.

Displaying the whole body of a digital witness in life-size underpins multiple consequential aspects of the user experience. It facilitates the realism [AG21], immersiveness [AG19; BG20], and naturalness [Art+14] of the interaction. In addition to exhibiting additional as well as more pronounced non-verbal cues, life-size whole-body reproductions of witnesses are perceived as more present than, e.g., displays of talking heads on a small screen [Brü18; BBG19]. The lifelike visual embodiment promotes users feeling co-present and as if sharing the physical space with the digital witness [AG21; Gam21; Ale21]. Feeling co-present amplifies the cognitive and affective connection of the user with the IDT due to a more engaging and immersive interaction [Fro18; AG21]. Users also feel physically, psychologically, as well as emotionally closer to the digital witness [Bal21b; BD21]. However, the digital witnesses might not appear as if they are in the same time frame as users [Kan14]. Dated clothing style, statements, or verbiage can reveal that the output of the IDT is a recording of the past [BBG19] and not contemporary with the user. However, we encountered no rationale on whether this illusion of contemporaneity is actually required or desired.

IDTs present users with a sequence of respectively disjunct recordings, i.e., their beginnings and endings do not fully align. Transitions between IDT outputs can be abrupt and noticeable, which subverts the perceived naturalness and plausibility of the displayed behavior of the digital witness. Figure 2.3 depicts the three types of transitions if IDT answers can not be interrupted. In the absence of user input, IDTs continually loop recordings of the idle witness to reflect their waiting state. Upon reaching the end of this recording, the output transitions to the beginning of the next idle recording; usually, the same idle recording is simply repeated over and over again (a). At any point during the idle loop, user input can prompt displaying an answer. Consequently, transitions from any point of an idle recording to the beginning of any answer can occur (b). Lastly, upon finishing an answer, the output transitions from the end of the recorded answer to the beginning of an idle recording (c). These lead to a wide range of potentially abrupt transitions. If IDTs accept and react to user input while already displaying an answer, transitions from any point of any recorded answer to the beginning of any recorded answer are additionally possible. To address all these cases, transition effects are used. An initial approach considered morphing between successive videos, but was found lacking and discontinued [Art+14] in favor of dissolve effects [Tra+15b; BBG19].

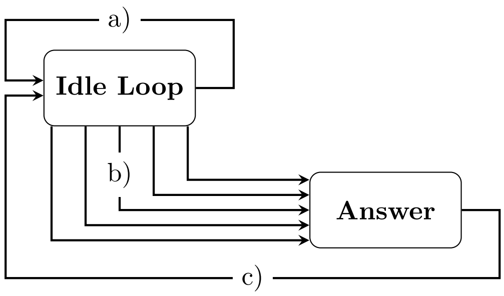



**Figure 2.3:** The three possible transitions between IDT outputs, if interrupting IDT answers is disallowed: a) End of the idle loop to the beginning of the next iteration of the idle loop, b) Anywhere during the idle loop to the beginning of an answer, c) End of an answer to the beginning of the idle loop.

Prior research considered audio-visual output modalities, with a focus on visual qualities. State-of-the-art IDTs are not limited to a single type of visual display technique: 2D, Pepper's Ghost, or stereoscopic 3D are all commonly used. However, we found no empirical user studies investigating possible differences in the effect of different display methods.

### 2.2.5 Conversational Ability

While IDTs aim to replicate conversations with real humans, their conversational ability is not on par with a real human being. IDTs are required to self-sufficiently handle the disparity between expected and actual conversational ability as well as associated shortcomings. The limitations arise from the restriction to provide exclusively pre-recorded reactions as well as the qualities of the NLP. The number of recorded responses is invariably finite, which limits the amount of afforded topics and matching replies [Art+15; Gam21]. Additionally, recordings are fixed and immutable, which means that IDT responses can not adapt to address current issues [Brü18; BDG21] or to the way the users phrased their questions [Dud21]. In practice, users who phrase their questions naturally or conversationally face difficulties receiving desired information from IDTs [Sch21; Hei21; Gam21], since semantically or syntactically complex input overwhelms their NLP. Consequently, users may be required to adopt a less natural or intuitive way of asking. To mitigate issues, implementations frequently necessitate additional assistance, such as instructions, example questions, or human facilitators [Glo21; Mar+22; Bal22]. IDTs that can only handle simple input are treated more like virtual assistants than virtual humans and elicit less empathy [Sch21]. Dissatisfied users who believe that the response of the IDT does not fit their question readily interrupt and ask again [Glo21].

Robust IDTs need to be able to conversationally handle numerous different types of irregularities or exceptions. These include, but are not limited to, off-topic questions [Art+15], input errors [BDG21], technical issues [Gam21], and input stemming from misconceptions [SH21] or knowledge gaps [Tra16]. IDTs employ fallback responses of varying genericity that are displayed if user input is detected and no suitable answer is found [Tra+15b]. These can disclose the technical limitations ("I'm actually a recording, so I can't answer that question.") [Gam21], direct the user elsewhere (transl. "I was not asked a question about this, but perhaps the following resource can help to answer your question.") [BBG19], attempt to steer the conversation toward an available topic ("Why don't you ask me a question about Auschwitz?") [Gam21], or offer advice for troubleshooting ("Maybe you should try to reboot.") [Ale21; Gam21]. However, even IDTs with multi-stage fallback designs [Tra+15a] do not identify or distinguish the reasons why no suitable answer was found. Consequently, fallback replies are indiscriminately selected without any relation to the cause of the irregularities or exceptions.

Affording multiple variations of the same recording [Tra+15b] could make interactions with IDTs less predictable and machine-like, as well as allow adapting the level of detail to the user [Mar+22]. Recording multiple versions for the same prompt, however, entails recording fewer answers to different questions. Additionally, IDTs were not equipped to detect or categorize users, which is a prerequisite for selecting specific answer variations for specific audiences.

### 2.2.6 Future-proof

IDTs need to adapt to time-related changes to remain viable in the future, which necessitates dedicated arrangements and precautions. The input processing system requires continuous revisions to reflect the linguistic changes and developments over time [MCW15], e.g., if the ASR fails to recognize non-binary gender forms [SU21]. As the content of IDTs is pre-recorded, it contains cues about its dated nature [Kan14], which may become more noticeable over time. However, IDTs usually do not attempt to hide their dated nature, e.g., by explicitly mentioning the date of recording [BBG19]. Consequently, future-proofing does not entail simulating social simultaneity with users, like using computer-generated data to continuously update the clothing style or choice of words of the recorded witness.

Instead, IDTs need to prepare for future technological advancements and requirements. The available display methods are limited by the recorded qualities [BBG19; BG20; BDG21], such as captured modalities or resolutions. To support a broad range of current as well as potential future displays [Art+14; Tra+15b], the recording process aims to capture as many dimensions as possible in as high quality as possible [Tra16]. Consequently, the recordings have unused qualities that intentionally exceed the capabilities of commonly prevailing displays [Art+14]. For example, IDTs recorded in 6K HD (5568×3132) are presented in 1080p (1920×1080 px) [BBG19] or displayed in 2D, although captured in stereoscopic or volumetric 3D [Gam21]. This allows IDTs to adapt and continue to appear modern as advanced displays become more common [Gam21].

Other IDT components face fewer limitations concerning adapting to the passage of time and can benefit more readily from technological advancements. Routine upgrades of outdated elements and sustained training of the NLP contribute to improving its future accuracy and processing time [Gam21].

IDTs need to be flexible and modular to adapt to changing requirements and conditions over time. The future possible display methods depend on the qualities of the recordings. Updates to the display method presuppose that the original recordings can be reliably retrieved in the future. Therefore, it is necessary that the recorded data can be stored safely, securely, and for an indefinite period of time.

### 2.2.7 Accessibility

IDTs aim to extend access to conversations with contemporary witnesses beyond the lifetime of the witnesses [BG20; Ale21]. The need to improve accessibility for a broader user group was recognized early [Art+14], but has not been addressed since.

IDTs designed for a single language lead to language barriers, e.g., when exhibiting IDTs requiring English-language input in a country where English is not a national language [Glo21; Bal22]. Additionally, ASRs work differently well for different user groups [Ale21], depending on the underlying sets of training data. Gender, pitch, intonation, speaking volume, accent, and dialect have been found to restrict the success of speech recognition of current IDTs [Bal21a; Mar+22; Bal22]. IDTs that require - in addition to natural language - further specific user input (e.g., push-to-talk), disadvantage people who struggle with the input design [BDG21].

In many cases, access to IDTs is limited to dedicated sites, such as museums. The considerable hardware requirements of cutting-edge IDTs can restrict the number of potential sites of operation [MCW17]. This can be counteracted by creating and providing additional, less demanding versions that are compatible with more affordable and available systems. Ethical concerns may prevent public online access to IDTs [Mar+22]. Due to the closing of museums during the SARS-CoV-2 pandemic, however, online access to some IDTs was made possible [Gam21].

While some barriers to use are mentioned, prior work does not address the range and diversity of user needs or preferences. In our review, we encountered no approaches to support or accommodate disabilities and impairments. We identified the need for IDTs to allow different ways of access.

### 2.2.8 Authenticity and Integrity

Authenticity and integrity are key characteristics and expectations when it comes to reliably reproducing encounters with real people [BG20]. The authenticity of the IDTs originates in the exclusive use of recordings of genuine testimonies and reactions of real people [For22]. Changes to the IDT, such as editing the recorded material or modifying the way pre-recorded answers are matched to user questions, can challenge its integrity and authenticity. This also means that IDTs are not compatible with generative AI applications such as Large Language Models [Vas+17; Nav+23], as the output generated by the latter commonly differs syntactically or even semantically from the original statements of the witnesses.

Yet, editing can not be avoided entirely. For example, extracting individual response clips requires processing and cutting the recorded raw material [BBG19]. It is important that the response is not manipulated, censored, or altered in a way that changes the original meaning [MCW17; Gam21]. However, the decision on exactly which parts to include and which to exclude is not definite. For instance, while silent pauses for thought before answering do not contain verbal information, they may contain body movements and other non-verbal cues that are meaningful parts of the complete response [Tra+15b]. Other editing dilemmas revolve around tailoring clip lengths to user groups, removing factual inaccuracies, or extracting answers that are contained in other answers [Mar+22]. Uncertainty and subjectivity are also inherent to the IDT's matching system, whose training relies on intervention by humans who evaluate the correctness and the authenticity of a match [SU21; Dud21].

In order for all these processing steps and decisions to be traceable and for the authenticity and integrity of the result to be verifiable, the unaltered raw material must remain available in its entirety. This reiterates our prior conclusion (see Future-proof) that it is necessary that the recorded data can be stored safely, securely, and for an indefinite period of time.

## 2.3 Requirements arising from Literature

Based on our literature analysis, we identified six requirements for IDTs. To apply to both existing and potential future IDT designs, we chose to use abstract rather than quantitative or binary requirements. The requirements for IDTs can influence, interact with, and conflict with each other, which adds further prohibitive complexity to their quantifiability and measurability. For example, an IDT using specialized displays to produce a more lifelike output might, in turn, become less accessible. Ultimately, the requirements represent ideals that implementations can and should strive for but are unlikely to realize fully.

- **R1: Authenticity:** The content and reactions of IDTs need to be authentic, i.e., recordings of actual statements by a contemporary witness instead of synthetic or AI-generated reactions. Additional arrangements may be required to allow the authenticity of an IDT to be verifiable.
- **R2: Conversational Input:** IDTs need to afford natural conversational user input. This includes robust and timely recognition and processing so that each input can be matched with an appropriate output.

- **R3: Lifelike Output:** The digital representation of the contemporary witness should be displayed in such a way that it appears plausibly realistic and natural to users.
- **R4: Accessibility:** To facilitate comprehensive and equal accessibility, IDTs need to support diverse input and output methods (e.g., different languages) to meet user needs.
- **R5: Revisability:** To continuously extend and improve usability, IDTs need to afford the ability to monitor, review, and adjust their system behavior. Additionally, IDTs need to provide for compatibility with and integration of future technologies.
- **R6: Self-sufficiency:** Users should not have to rely on intermediaries or other types of external support to have meaningful interactions with IDTs.

## 2.4 Previous IDT Design Models

We identified two previous models for IDT designs, one ([Tra+15b]) originating from the project Dimensions in Testimony and one ([MCW17]) from the Forever Project. Neither model fulfills all requirements listed in the previous subsection (see Table 2.1).

**Table 2.1:** Overview of which requirements are met by which previous model. ✓ indicates fulfilled requirements, ✗ indicates unfulfilled requirements, and ✓/✗ indicates requirements that are addressed, but only partially fulfilled.

| Requirement | [Tra+15b] | [MCW17] |
|---|---|---|
| R1: Authenticity | ✓/✗ | ✓/✗ |
| R2: Conversational Input | ✓ | ✓ |
| R3: Lifelike Output | ✓ | ✓/✗ |
| R4: Accessibility | ✗ | ✗ |
| R5: Revisability | ✓/✗ | ✓/✗ |
| R6: Self-sufficiency | ✗ | ✗ |

In [Tra+15b], Traum et al. detail their approach to preserving conversations with Holocaust survivors (see Figure 2.4). By using exclusively pre-recordings of Holocaust survivors, the proposal fulfills the first aspect of R1: Authenticity. However, the model lacks clear means or provisions on how the authenticity of the output material can independently be examined or verified.

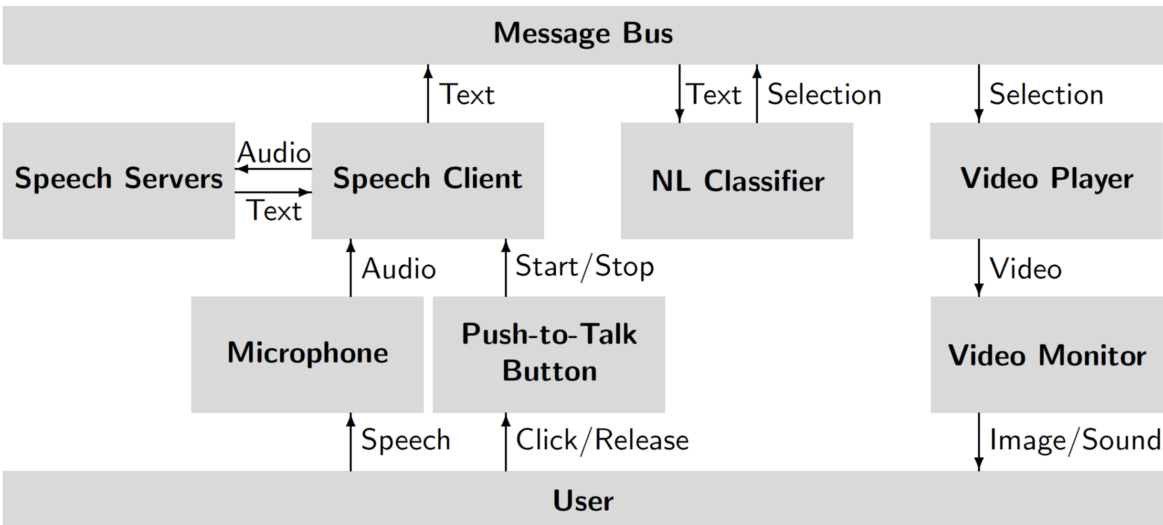



**Figure 2.4:** IDT architecture of the *Dimensions in Testimony* project as defined in [Tra+15b]. Reproduced with permission from Springer Nature.

A microphone, which can be triggered by pushing and holding a button, allows user input in spoken form. The recorded audio is converted to text and subsequently used by a natural language classifier to select an appropriate response. While the corresponding implementation described in the paper showed lackluster accuracy of the matching system, the features of the theoretical model satisfy R2: Conversational Input. The model also fulfills R3: Lifelike Output by incorporating a high-quality audio-visual display, transitions between the displayed recordings, and idle loops between IDT responses. However, it neglects R4: Accessibility by not affording alternative or customizable methods for input or output. Additionally, while the paper mentions continuous training of the input classification system and differences between recorded and employed video quality, it contains no explicit explanation of how the original raw data is to be used or managed, leading to but partial fulfillment of R5: Revisability. Finally, the paper acknowledges the reliance on human facilitators for user interactions with the system, which was however omitted in the corresponding diagram. Consequently, the described approach does not satisfy R6: Self-sufficiency.

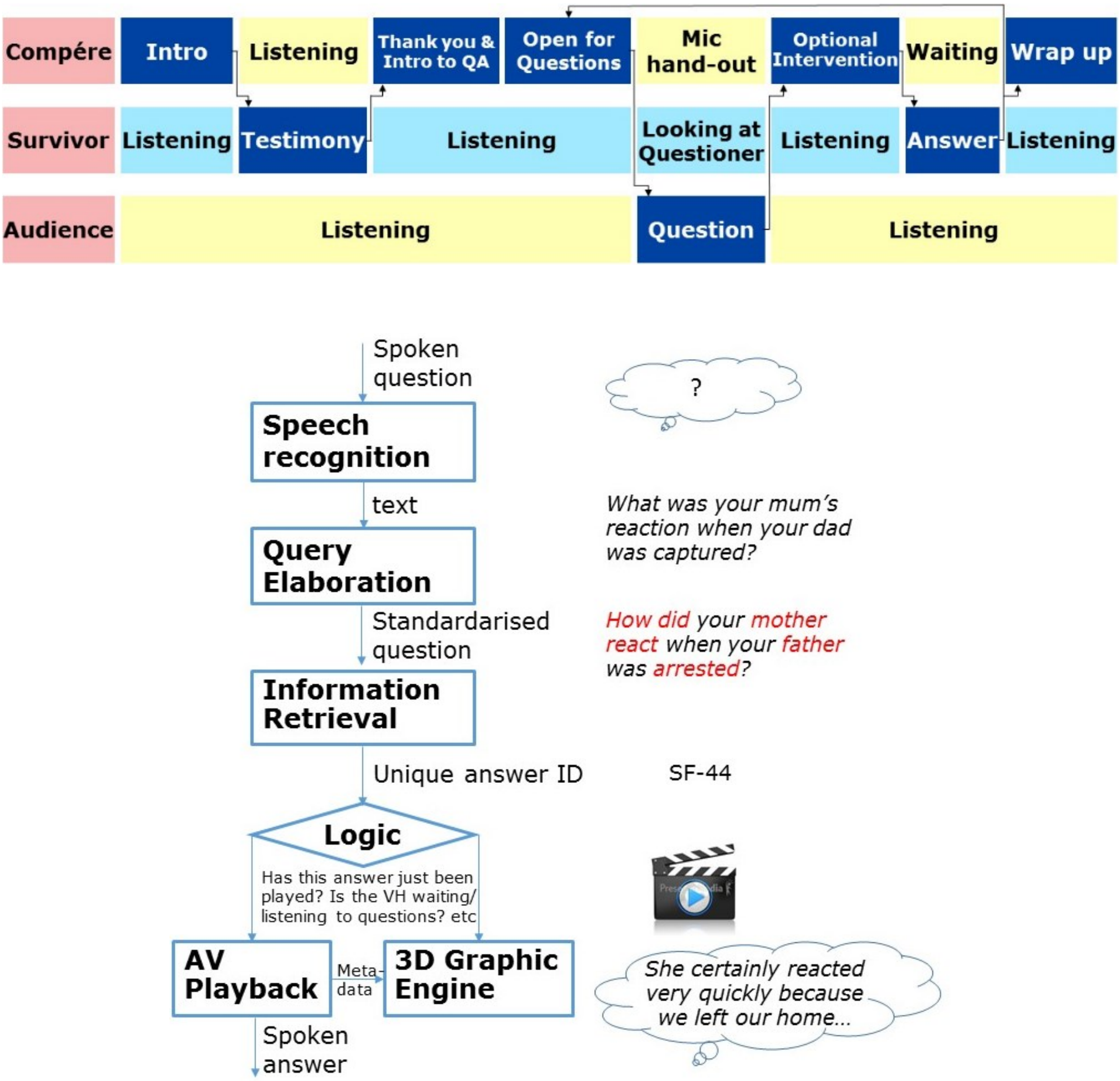



**Figure 2.5:** IDT architecture of the *Forever Project* as defined in [MCW17]. Reproduced with permission from Springer Nature.

The model by Ma et al. for the Forever Project [MCW17] (see Figure 2.5) shares several aspects with the model by [Tra+15b]. The main difference lies in the use of computer-generated images of the contemporary witness (see Figure 2.6) to provide a more personalized and immersive experience while the system waits for user input. However, due to the Uncanny Valley Effect, this comes at the expense of displaying consistently natural behavior. Both the computer-generated and the pre-recorded portions of the digital witness are presented using a high-resolution stereoscopic 3D audio-visual display and transitions. Consequently, this model partially fulfills R3: Lifelike Output. Due to the use of computer-generated images of the contemporary witness and the absence of explicit arrangements for verifying the authenticity of the presented statements and reactions, R1: Authenticity is likewise only partially fulfilled. A button-activated microphone enables users to input information through spoken language. The captured audio is then converted into text and utilized by a statistical relevance model to promptly determine the most suitable response. The model therefore meets R2: Conversational Input. However, the model does not satisfy R4: Accessibility, as apart from the aforementioned devices and modalities, no alternatives for input or output are planned or provided. While the paper mentions recording additional data for future use, the accompanying model considers its components and devices to be static and offers no method for their upgrade or replacement. This combination results in partial fulfillment of R5: Revisability. Lastly, the paper recognizes that "without the facilitator, the interaction does not work well" [MCW17]. Therefore, the outlined model fails to meet R6: Self-sufficiency.

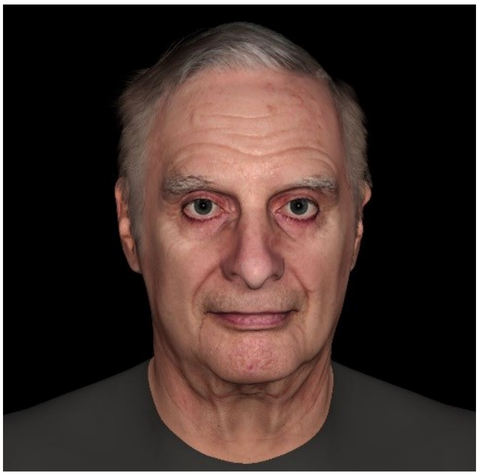


**Figure 2.6:** Computer-created image of the contemporary witness as shown in [MCW17]. Reproduced with permission from Springer Nature.

## 2.5 Summary

In this chapter, we provided first partial answers for **SQ-1:** "What are the requirements to digitally preserve interactive conversations with contemporary witnesses?" and **SQ-2:** "What are the associated practices or shortcomings of current IDT implementations?".

We compiled, reviewed, and analyzed literature on IDTs to examine the current state-of-the-art of IDTs. Building upon our analysis, we identified that Authenticity, Conversational Input, Lifelike Output, Accessibility, Revisability, and Self-sufficiency are requirements for IDTs. We subsequently investigated the extent to which existing IDT design models meet these requirements.

# Chapter 3

# Expert Interviews

*"Kunst = Natur - x"*
*[Art = Nature - x]*

— Arno Holz, *1891* [11]

## Contents



In this chapter, we describe how IDTs are provided and used, based on the use case of Holocaust survivors. We explore characteristics of current IDTs that impede their effective application as well as unfulfilled potentials. Despite the effortful creation and growing application of numerous IDTs over recent years, there is little published empirical research on current IDTs, their implementations, and their effects. To bridge this gap in research, we interviewed 26 IDT experts across six roles involved with creating, operating, using, or investigating IDTs. We asked our interviewees about their experiences and observations, including success stories and challenges during their involvement.

In Section 3.1, we detail our chosen study design, the procedure, the composition of the user sample, and our method for analyzing the gathered data. We then discuss our findings in Section 3.2, explain their limitations in Section 3.3, and use the gained insights to amend the list of requirements in Section 3.4. We summarize this chapter in Section 3.5.

## 3.1 Method

For a more comprehensive understanding of the current state-of-the-art of IDTs of Holocaust survivors, we aimed to capture the diverse perspectives and extensive experiences of the people involved. Consequently, we recruited broadly across the roles concerned with creating, operating, using, or evaluating these IDTs, e.g., Holocaust survivors, their bereaved, domain scientists, user experience

[11] [Hol91]

researchers, designers, and museum staff. Our selection criteria focused on participants with comprehensive professional experiences with IDTs, such as paid or voluntary occupational backgrounds. As a consequence of their backgrounds, each participant had numerous hours (frequently hundreds or even thousands over multiple years) of first-hand experience as expert user as well as vast second-hand experiences through observation.

We used both convenience and snowball sampling [Hec11] to reach potential participants from this highly specialized and thus narrow target group. We gathered our initial set of volunteers by contacting our pre-existing personal connections in this field. Additionally, we arranged for visits to educational institutions exhibiting IDTs and approached the experts on-site. At each location, we recruited multiple participants as well as different roles when possible.

The study design was individually reviewed and approved by our institution's ethics committee and its data protection officers to protect the well-being and rights of our participants. Particularly, the inclusion of Holocaust survivors required further sensitivity [Mas+10] and consideration.

### 3.1.1 Study Procedure

Before the interview, we provided each participant with a consent form informing them of the study procedure, how their data would be recorded, processed, and stored, as well as their corresponding rights. We also explained our use of the term IDT and offered to adopt a different term during the interview if preferred by the interviewee. If the participant provided their informed consent, we started the recording and conducted a semi-structured interview. We began the interview by asking the participant to detail the extent of their involvement and experience with IDTs. We followed up with questions about the purpose, functionality, and properties of IDTs. In the main part of the interview, we asked the participant to walk us through a typical instance of their involvement with IDTs (e.g., design, production, use), share satisfactory characteristics, and reflect on any flaws they encountered. The last group of questions focused on what the participant would improve or change about IDTs. The full semi-structured questionnaire can be found in Section A.1 of the Appendix. At the end of the interview, we summarized our notes to give the interviewee the opportunity to amend our understanding or append statements. Afterward, we stopped the recording, wrapped up the interview, and asked the interviewees for recommendations for potential further study participants.

### 3.1.2 Participants

We conducted 26 interviews in total, 20 in English and six in German. 18 interviews took place in-person, while eight were carried out remotely via video calls with Zoom[12]. All interviews were conducted by the same interviewer (author of this thesis) for internal validity and consistency. Our participants included but were not limited to IDT-suppliers, staff of six different Holocaust museums, as well as independent, unaffiliated experts. Each participant (see Table 3.1) had one or multiple of the following roles:

- **Contemporary witness (CW):** Holocaust survivors of whom at least one IDT was created.
- **Facilitator (FA):** Guides and observers of large numbers of other users and simultaneously expert users themselves.
- **Management (MM):** Administrative staff responsible for IDT exhibitions, public relations, funding, or similar.
- **Bereaved (BR):** Users who interacted with the IDT of a late contemporary witness who was their family member or close friend.
- **Producer (PD):** Employees involved in the conception, development, and operation of an IDT.
- **Researcher (RE):** Domain experts and user experience researchers with a broad understanding and overview of IDTs and simultaneously expert users themselves.

[12] https://zoom.us/

**Table 3.1:** Participants, their corresponding roles, and levels of experience at the time of the interview. * marks intermittent levels of experience.

| ID | Roles | Experience | ID | Roles | Experience |
|---|---|---|---|---|---|
| P1 | PD, FA, RE | 3 years | P14 | CW | 1 year |
| P2 | CW | 4 years | P15 | RE | 5 years |
| P3 | FA | 6 years | P16 | CW | 7 years |
| P4 | FA | 6 years | P17 | BR | 6 years |
| P5 | FA, MM | 5 years | P18 | MM | <1 year* |
| P6 | MM | 1 year | P19 | MM | <1 year* |
| P7 | FA | 1 year | P20 | PD | 13 years |
| P8 | FA | 3 years | P21 | PD, BR, RE | 13 years |
| P9 | FA | 3 years | P22 | PD, FA, RE | 4 years |
| P10 | FA | <1 year | P23 | PD | 9 years |
| P11 | FA, MM | 3 years | P24 | PD, FA, RE | 3 years |
| P12 | MM | 5 years | P25 | PD, MM, RE | 4 years |
| P13 | MM | 5 years | P26 | CW | 2 years |

With the exception of P18 and P19, all participants were actively and continually working with IDTs at the time of the interview. The experiences of P18 and P19 took place five years before their interviews.

### 3.1.3 Data Analysis

The audio recordings of the 26 interviews amount to 32.66 hours in total ($M = 1.26h, SD = 0.51h$). The shortest interview lasted half an hour, while the longest interview surpassed two hours. An extended overview of interview details can be seen in Table A.1 of the Appendix. In preparation for our qualitative analysis, we transcribed and anonymized[13] each recording and translated the transcripts of the German language interviews into English. In our analysis, we also included provided artifacts like museum handouts as well as our observations on location.

Following a bottom-up approach, the author of this thesis manually inductively coded the data, focusing on semantic meaning and an experiential and constructionist interpretation [Fli18] of the interviewees' statements. We used affinity diagramming [Scu97; HB98], which is a commonly used method in HCI research for analyzing interview and observation data [Gol+22; HH15; JL14], to iteratively consolidate these coded data points into high-level themes and sub-themes with shared meaning-based patterns related to the purpose, challenges, or effects of IDTs. During our four iterations of the analysis, we continually revised and refined our codes and themes.

### 3.1.4 Positionality

Our research approach is also built on our own perspectives and experiences. The German school education of the author of this thesis emphasized the need for historical responsibility and the dangers of historical negationism. This was reinforced in everyday lives by encounters with public memorials on the one hand and witnessing recurring instances of racial discrimination in general and anti-semitism in particular on the other hand. However, we are not part of a Jewish or Holocaust survivor community ourselves, and our academic backgrounds primarily involve disciplines of computer science instead of Holocaust education. Nevertheless, our target group welcomed and appreciated our research project as well as our technology-oriented perspectives. Additionally, the experiences of the author of this thesis working on and researching IDTs benefited us in establishing rapport as well as empathizing with interviewees.

[13] P21 explicitly requested to be de-anonymized after the analysis and for their identity as Storyfile's Stephen D. Smith to be revealed.

## 3.2 Findings and Discussion

In our analysis, we identified three high-level themes:

- Shortcomings during interactions with IDTs
- Opportunities and restrictions regarding access and availability of IDTs
- Instances of unanticipated emotional functions of IDTs

Figure 3.1 shows the accompanying thematic map of our analysis. Within each high-level theme, we detail and discuss our findings in the corresponding sub-themes. At the end of each high-level theme, we present implications and opportunities for improved designs. During our analysis, coded insights could be relevant and assigned to more than one theme. Consequently, our themes are not strictly disjoint and intersect on some features.

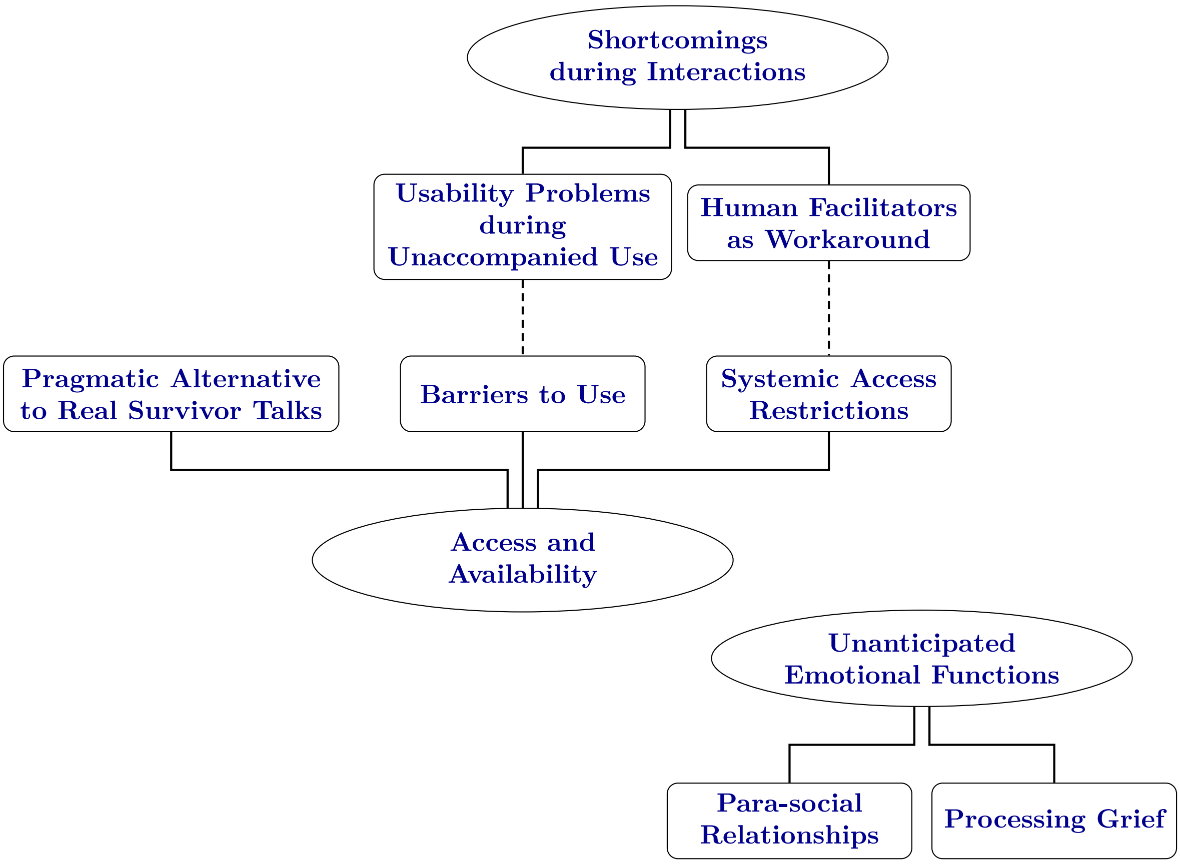



**Figure 3.1:** The themes and sub-themes derived from our analysis of expert interviews on how IDTs of Holocaust survivors are provided and used.

### 3.2.1 Shortcomings during Interactions

In this section, we present the issues users faced during unaccompanied interactions with IDTs and how accompaniment by human facilitators aimed to solve these issues. We detail the different categories of facilitation and their respective drawbacks.

#### 3.2.1.1 Usability Problems during Unaccompanied Use

Our participants reported that users commonly overestimated the intuitiveness and ease of interactions with IDTs, partly due to expectations of conversational abilities on par with a real human being. However, IDTs were unable to cover all topics as their pool of available answers contained exclusively pre-recorded responses. Likewise, user questions on topical events or other developments that took place after the recording could not be addressed by the IDT (P3, P5, P15, P21, P24, P25, P26). These gaps in content were inherent to the underlying concept of IDTs [Kan14] and were both expected and accepted by our interviewees.

Several participants mentioned issues that were related to the input design [BDG21]. Complex syntax, nested questions, or flowery phrases overburdened the NLP, which then failed to identify the intent of the verbal user input and return a matching response (P1, P3, P4, P6, P7, P8, P10, P12, P19, P24, P25). While a mismatched response could still be interesting, it broke immersion and deprived users of the rewarding feeling after a successful interaction. Repeatedly receiving unfitting replies from the IDT even left users dissatisfied or annoyed: *"Then restlessness spreads somewhat, there is some laughter, or frustration becomes noticeable" (P24).* Users frequently expected greater flexibility in formulating questions and found it difficult to adopt less natural, simpler sentence structures. The IDT implementations also required users to provide additional input signaling the beginning and end of their questions, e.g., by utilizing push-to-talk. Several participants (P1, P7, P10, P11, P19, P22, P23, P24) reported usability problems stemming from the increase in cognitive load, especially for people unaccustomed with these interaction methods: *"How to ask a question is not intuitive because you still have to, it's like a press and hold. [...] It's not like I can say 'Hi!' and [the IDT] knows and responds" (P23).*

Our interviewees also observed reluctance to interact at all due to uncertainty about what questions to ask (P4, P8, P9, P11, P18, P22, P24). P18 noted that particularly in group settings *"there was some sort of hesitation. 'I don't wanna be the first one or the only one to ask' "*. Users were too shy or anxious to engage with the IDT out of fear of embarrassing themselves by asking a question that might be perceived as *"nonsense"* (P22). If not provided with an introductory testimony, the lack of knowledge of history or the life of the witness further inhibited the ability of users to think of specific questions (P7, P8, P23).

During applied use, our participants encountered various technical issues and glitches within any IDT component, including Internet connection, input devices, ASR, and output devices. However, the overall shortage of system feedback or diagnostics was a recurrent cause of uncertainty or confusion. If the IDT reacted in an unexpected manner, it was difficult to identify whether an error occurred while the user operated the input device or during the processing of the user input (P1, P21, P22). Since the pre-recorded responses did not always align with the precise phrasing of the user question, distinguishing between a non-matching answer and a matching answer that was enclosed by another, longer response was challenging (P1, P5, P7, P8, P10, P21, P22, P24). Similarly, receiving a fallback answer left users with residual doubts about whether, in fact, no matching answer exists within the IDT's database of recordings. P24 pointed out that current IDT designs attempted to strictly reproduce the aspects of physical survivor talks and possibly underutilized their potential as a digital medium to address shortcomings: *"It seems to me like a form of skeuomorphism, that you somehow imitate a thing from before so much that you limit yourself in a way that is not productive at all."*

#### 3.2.1.2 Human Facilitators as Workaround

To address these usability problems, many sites chose to have employees or volunteers accompany and support users. While this conflicted with the initial objective of immediate and independent interactions with IDTs, the issues were deemed severe and detrimental enough to warrant a pragmatic trade-off.

Our participants reported that the main responsibility of facilitators was to make it easier for users to receive a satisfactory response. We encountered different implementations of this responsibility (see Figure 3.2), which confirms and extends prior findings [Glo21; Bal22]. In our analysis, we constructed two categories: Guided facilitation, which aimed to provide a predefined, structured format for IDT interactions, and unguided facilitation, which featured fewer instructions and restrictions but also less proactive support.

Guided facilitation sessions began with an explanation of the system and the session procedure to an audience of users (P5, P8, P10, P11, P18, P23, P24). This openness about the technical and non-human nature of IDTs also helped set user expectations at an appropriate level *"'cause we're not trying to trick you. We want you to understand you're interacting with a video and AI technology, but that is a real person's story that you're interacting with" (P23).* After displaying a pre-recorded introductory testimony, the audience members were encouraged to ask questions. Frequently, the facilitators themselves commenced the interactive section of a session by asking the IDT light-hearted icebreaker questions to demonstrate the functionality and help users overcome their shyness (P8, P10, P11, P17, P24).

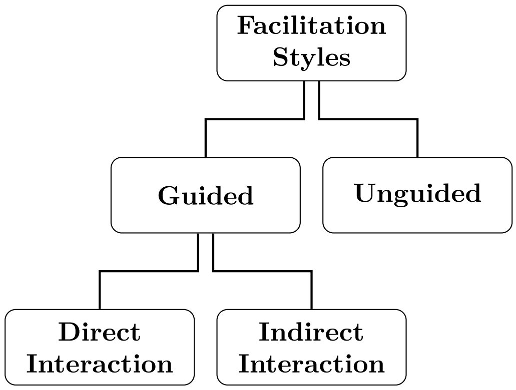



**Figure 3.2:** Overview of the different facilitation styles we encountered during this study.

Since the session duration was predetermined by scheduling and staff availability, museum visitors and other users were limited in the number of questions and topics that could be covered (P5, P8, P9, P10, P15). Usually, the overall length of an entire session was less than 60 minutes.

Within the category of guided facilitation, we identified two subcategories with regard to handling user input: Direct interaction and indirect interaction. During direct interaction facilitation, users were allowed and invited to ask their own questions themselves. Consequently, facilitators in this subcategory also explained the interaction design to prospective users. If users faced difficulties when asking the IDT a question during the interactive section of a session, the facilitators encouraged them to try again, gave pointers, or ultimately offered to ask in the user's stead (P5, P7, P8, P9, P23, P24, P25). In many cases, users received pre-selected and pre-formulated sample questions as additional guidance (for example, in the form of handouts), which, however, also biased their input selection.

During indirect interaction facilitation, IDTs received input exclusively from facilitators, who sourced and selected suggestions for questions from the audience (P3, P4). Building on their prior experience and familiarity with the matching system, the facilitators then attempted to rephrase and input the question in the audience member's stead so that the IDT would return a fitting reply. However, this made users rather observers of a simulated conversation between an IDT and a facilitator and gave users the least agency. For example, facilitator P3 described omitting or substituting audience questions if they considered the question not relevant: *"Sometimes I'll steer it a little into a different direction so we get the important content of the [IDT]. You know, without being too obvious that you're not appreciating the question that was just asked."*

Despite the encouragement and support by facilitators, both subcategories of guided facilitation experienced instances without any questions from the audience (P4, P9). In these situations, the respective facilitators chose their own questions to fill the interactive section: *"Sometimes the visitors are very silent and do not say anything at all. Even when I say, would you like to ask a question? So, in that case, my response is that I just ask all the questions. But it happens quite frequently" (P9).* However, no interactions by users and consequently no effects of interactivity on learning or engagement [Fro18; Mar+17] occur.

Unguided facilitation used no rigid framework or session-based access to IDTs. *"I'm here to help you or assist you, there's no timed session, there's no scheduled things. It's really just come in at your leisure and use it as you like" (P11).* It allowed users to interact with the IDT directly and ask their own questions. The respective facilitators were situated nearby and provided support if requested by users. However, P11 also acknowledged that many interactions by visitors were short and limited to one or two questions before moving on. Another implementation used unobtrusive facilitators as silent input detectors who observed the audience and pressed the push-to-talk button if they noticed user behavior indicating an upcoming question (P7). This approach bypassed issues related to users themselves having to signal the beginning and end of their questions. In addition to the different extents of guidance and allowed direct interaction during facilitation, some participants also reported inconsistency between facilitators, which added further variation to user experience and agency (P8, P10, P15, P18). Museum staffers like P10 put the focus of their facilitation on the IDT itself: *"Every facilitator has their own style and my style is very much trying to take myself out of it as much as possible. So I don't do an additional bio. I just kind of say, 'This is what's gonna happen. They're gonna tell you about themselves and then we will do the question asking' and I'll explain how that works and I just try to make it about them as much as possible."* Other facilitators adopted a more active role, favoring a stronger connection between user and facilitator (P4, P9). This variation was also accompanied by different levels of background information and technical details presented to users. Qualification and training required of facilitators consisted mostly of extensive time spent interacting and familiarizing themselves with the IDTs available (P5, P7, P8, P9, P11, P18). Since their capabilities depended on the extent of their practical experience, the more seasoned facilitators commonly volunteered to conduct more challenging sessions, e.g., if unruly behavior by museum visitors was expected (P8).

Our participants described that facilitators were also commonly tasked with the technical operation and maintenance of IDTs (P5, P9, P10, P11, P18, P23, P24). While facilitators accumulated technical knowledge through experience, they usually were afforded little more tools or means to solve a given problem. Effective approaches centered on generic solutions, such as rebooting the system (P9, P11) or using their private smartphone and data to bridge an Internet outage (P24). When faced with more serious or persistent technical issues, they had to rely on their external IDT provider for support: *"When it's like just beyond our control where we'll have to get a person who's literally not in the state at all to come to try and fix it for us" (P9).*

#### 3.2.1.3 Discussion and Implications

**Make IDTs more self-sufficient.** Our analysis revealed that human facilitators are not able to thoroughly solve all the usability problems and flaws of current IDTs. The issues frequently arose from a habitability gap [Moo17b], i.e., human-like presentation without human-like conversational ability [Art+15; Gam21]. On top of that, the use of facilitators introduces other problems and restrictions. The access and use of an IDT are limited by the availability of a human facilitator. Distinct inconsistencies between facilitators and facilitation methods result in varying quality and lack of reproducibility of user experiences. While facilitators positively provided assistance and helped users, they also biased or constricted users in their choice of questions and topics. Self-sufficient IDTs that eliminate the necessity for a human facilitator could remove their corresponding restrictions, thereby allowing for more agency of users and broader applicability of the system by means of independent, direct, and unmediated user interactions [Hei21; Art+15].

This means that users need to be supplied with more information and feedback during the interaction to afford them a more accurate mental model of the functionality of the IDT. Instead of receiving the same generic fallback response, users could be made aware of why they did not receive an answer and adjust their behavior and input accordingly. Explanatory fallback replies reflecting different conditions could provide diagnostics in an unobtrusive and conversational manner. For example, questions on topics that happened after the recording for the IDT could be met with a response mentioning the date of the recording, reinforcing the temporally past and technical nature of its content. Alternatively, an option for displaying diagnostic details like the output of the ASR or the original question to a matched IDT response could also help users comprehend the system.

Additionally, our investigation confirmed the need for an IDT-inherent introductory testimony to provide users with necessary context and knowledge [BBG19; Gam21; SU21; Hei21].

**Explore other input detection methods for IDTs.** Current IDT implementations require additional and distinct user cues for the automatic detection of the beginning and end of their questions. As the required behavior is not part of natural conversations, it is perceived as unintuitive and a frequent cause of input errors. IDT designs could explore alternative methods to detect natural cues preluding a user question. Combining multiple modalities [KB18], like posture, gesture [Pom+20] and gaze detection [Ver+00; McM+19], could result in a system that is both more robust and more intuitive to use.

**Offer IDTs that can be experienced by oneself.** Apart from sparse browser-based interactions on personal devices, we encountered no offers for users to experience and interact with IDTs alone. The interaction opportunities were accompanied and observed by facilitators and commonly other users as well. However, the continued encouragement by facilitators during guided sessions did not prevent recurring instances of users being too reluctant to ask their own questions. Shy or anxious users could benefit from unaccompanied, private interactions without the potentially unwanted attention and observation by other people. However, this would also require IDTs themselves to be more self-sufficient and approachable. This could include additional encouragement while the IDT waits for input, e.g., a more inviting idle loop [Gam21].

### 3.2.2 Access and Availability

In this section, we detail how IDTs aim to provide broader access to conversations with, e.g., Holocaust survivors and the corresponding challenges. We describe implicit or inadvertent barriers for user groups as well as systemic limitations to access and availability of IDTs.

#### 3.2.2.1 Pragmatic Alternative to Real Survivor Talks

Our participants unanimously stated that the main purpose of IDTs is extending the bounded and fleeting temporal availability of encounters with contemporary witnesses into the future, beyond their passing. Digital lifelike simulations of conversations with, e.g., Holocaust survivors aim to preserve access to an important component of history and civic education for future generations. As long as the opportunities remain, however, talking and interacting with real human witnesses was roundly regarded as superior and preferred: *"Obviously, it's always better when I go to a school and I tell my story and tell my experiences and answer questions; it's obviously much, much better. But at the same time, this cannot go on forever" (P16).*

Four participants (P20, P21, P24, P25) indicated that the aspired persistence of IDTs is challenged by the uncertainty of the future. While technology progresses steadily, precise advancements are difficult to predict. Over time, static designs and implementations will become obsolete, warranting revisions and novel re-implementations. New IDT iterations, however, will still need to draw from the same original recordings, as these can not be re-recorded. They consider technological progress as an important factor when deciding what qualities, like modalities or resolutions, are to be captured: *"So, and whatever comes up, you know, we don't know five years from now what's gonna come out. They should be able to use [the recordings]" (P20).* Consequently, current IDTs usually display only a fraction of the recorded qualities, which echoes our findings in Section 2.2.6.

Several participants (P1, P2, P5, P6, P7, P15, P16, P22) re-emphasized that IDTs are neither intended nor suited for replacing conversations with real human witnesses. While being aware of their imperfections and limitations, IDTs are nevertheless viewed as a viable, albeit lesser alternative. One Holocaust survivor assessed the IDT concept as *"from my point of view, it's plan B, but the plan B is almost as good. It's not as good, but it's almost as good as plan A: to have a real person standing in front of you" (P16).* Many of the same participants (P1, P5, P6, P10, P15, P16, P19) also argued for the pragmatic use of IDTs while the corresponding human contemporary witnesses are still alive and available. Even now, when one can still meet and ask Holocaust survivors, such opportunities are already limited. Additional digital alternatives could help to provide broader access to survivors' stories: *"Even the survivors that are still around have limited time and energy. Being able to ask their recordings is very useful so that they don't have to answer the same question a hundred times. So I think that's very useful, and also that it can travel around as it is a digital thing. It can be transferred much more easily than a human being can be at this stage" (P10).* The suspension of in-person survivor talks during COVID-19 reduced access further (P19). Subsequently, video calls with Holocaust survivors emerged as another digital alternative, which also removed the need to travel. The demand for IDTs remained, however, as Zoom calls with Holocaust survivors still could not scale sufficiently to provide an opportunity to every interested person.

#### 3.2.2.2 Barriers to Use

Initially, all of our participants described IDTs as accessible to anyone or free of restrictions. Museum exhibition manager P6 quipped *"basically everybody, and I joke with my staff, I say, you know, our target audience is people with skin".* After further reflection, however, many participants qualified their statements. The first necessary trait of prospective users is pre-existing curiosity about using IDTs (P1, P3, P4, P7, P9, P10, P11, P16, P21). They expressed that self-motivated and deliberate interactions are expedient for the purpose of conversations with contemporary witnesses. *"Ultimately, it's a prerequisite that people have an interest in using them" (P1)*, whereas compelling users to engage with IDTs would be counterproductive.

The second constraint listed by our participants is the cognitive or emotional maturity of the users (P1, P3, P14, P16, P17, P25). They proposed limiting access to IDTs to users above a certain age, although their respective age recommendations varied. While P3 could imagine allowing pre-school children to accompany their parents, most minimum age recommendations gravitated towards early and mid-teens: *"So I think that there should be a lower age limit, so 14 I think is good. It could also be twelve, depending on how prepared you are and how capable you are" (P25).*

The third implicit filter is the user's ability to understand and handle the respective interaction designs (P1, P2, P4, P5, P6, P8, P10, P11, P12, P13, P18, P22, P24, P25). Whereas the interaction designs of IDTs tend to benefit younger users, they impede older demographics. While IDTs strive towards providing natural interactions, their limited conversational ability often requires user input to be phrased in a specific, straightforward manner. Our participants reported that adjusting to the interaction came easier to younger users, who were also frequently more at ease due to tech-savviness or prior experiences with other CAs: *"The way that the digital interaction works, it's something that students are familiar with. They're familiar with technology and I think that they're just more comfortable interacting with it through this digital format" (P5).* Conversely, senior users are usually less familiar with the employed technology and consequently face more efforts in order to reconcile the interaction sequences with their mental models.

The fourth and final constraint is the ability to speak, hear, and understand the language of the respective IDT (P1, P5, P6, P8, P9, P10, P21, P22, P24, P25). This means that users with speech impediments, non-native speakers, or even native speakers with an accent or dialect are inherently inhibited from providing input that the IDT can process correctly. This issue stems from deficiencies in the underlying language models used to process natural language input as well as the limited selection of provided input modalities. Visual or hearing impairments additionally preclude users from receiving and perceiving the entirety of the IDT response, which is commonly exclusively audio-visual and without assistive features. Referring to the IDT of a female Holocaust survivor, P6 accentuated that *"unless you can hear what she's saying to you, you're not gonna be able to participate".*

#### 3.2.2.3 Systemic Access Restrictions

In many cases, the opportunity to interact with an IDT is limited to a small number of dedicated sites (see Figure 3.3 for the distribution in the contiguous United States as an example), which also utilize specialized hardware and a controlled environment. This stems from the significant and frequently prohibitive financial investments required for both, the production as well as the operation of an IDT (P1, P6, P9, P11, P12, P13, P17, P20, P21, P25, P26). Our participants reported experiencing difficulties while acquiring funding, be it from public grants or private donors, to cover the six-figure up-front costs for the creation of a new IDT. Since many interest groups can not afford this expense, it firmly restricts the overall number of IDTs in existence.

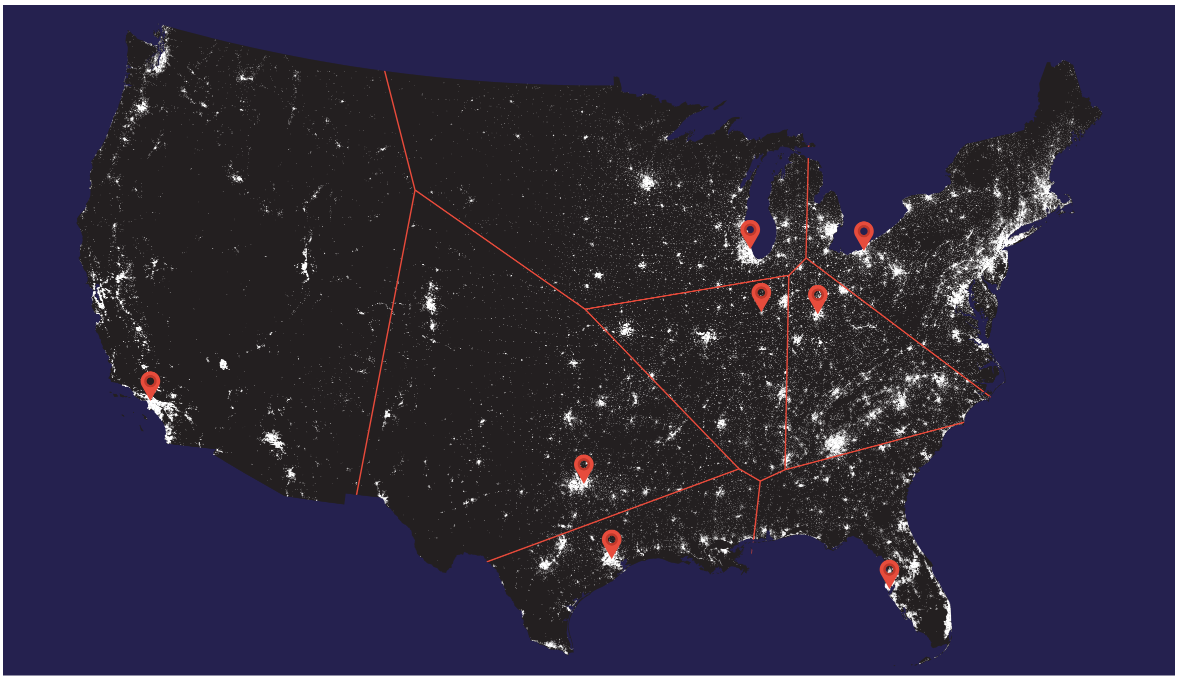


**Figure 3.3:** Distribution of sites providing in-person access to IDTs of Holocaust survivors in the contiguous United States. Each white dot represents 1000 people. An overlayed Voronoi diagram illustrates the respectively closest sites providing IDTs. This figure adapts the United States Census Bureau's 2020 *Population Distribution in the United States and Puerto Rico* map [Uni21].

Exhibiting an already-produced IDT is also accompanied by numerous expenses. In addition to hardware and staff, exhibition sites need to cover recurring licensing fees levied by the respective hosts of the IDT's back-end. Museum manager P12 disclosed that these licensing fees apply even for an IDT created through funding by their museum: *"From an administrative standpoint, budget, you know, it wasn't cheap. We still pay them a considerable amount of money, even though we don't technically own [this IDT]. Right. [The host] owns [this IDT], and we still have to pay them a quarterly sum, which is not cheap"*. Offering museum visitors a broader selection of the host's already existing IDTs, e.g., different IDTs on different days of the week, would incur additional licensing fees. Consequently, few institutions are able to persistently afford IDTs and even fewer institutions can offer access to more than a small subset of otherwise available IDTs.

Some participants (P1, P9, P10, P12, P15) further pointed out that restricted access to these institutions due to remoteness, opening times, or entry fees represent additional barriers for interested users. They reasoned that transportable versions of IDTs are essential for increasing the accessibility of IDTs: *"If you wanna reach a wide audience, you gotta bring it into schools. 'cause most schools in [my country] are not located close enough to a Holocaust museum to be able to go in and use it or have the funds to get students there" (P15).* The participants were aware that these alternative IDT implementations would likely require concessions and not provide the same deeply engaging experience as dedicated, immobile installations; a trade-off they considered worthwhile. Creating transportable implementations *"might mean not making it a fully immersive space, but that's okay because the important part is that it gets out to more people and it's shared" (P10).*

However, several participants (P5, P11, P12, P15, P20, P21, P22) also outlined the benefits of and the need for consolidated and controlled access to IDTs. They described the ability to monitor and intentionally restrict access to IDTs, as well as recorded material in general, as a critical safeguard against improper use, like editing or misconstruing recorded statements. Conflicting interests and one-sided dependencies arose, however, when the same institution is simultaneously responsible for providing access to IDTs and for deciding what constitutes improper use. On the operational side, P20 also acknowledged these ambiguities while navigating their authority and corresponding responsibility: *"We ultimately host everything so we can delete or not give anyone access to whatever, but you know, that's a gray area"*. A single centralized back-end, which is housed off-site, expedited maintaining and improving the IDT, e.g., by training the CA. With this single instance, no inconsistencies between instances could occur and synchronization measures were not required. However, the external hosting site, as well as the corresponding connection via the Internet, constituted critical single points of failure. P22 paraphrased the lack of transparency of distributed IDT implementations and their resulting inability to address such external issues on their own: *"We are dependent. So the testimony is not only on one device, but there are several instances involved and then also the server. And if there is an error somewhere, [...] something was changed in the API, and then suddenly no more videos were retrieved. So there are several sources of error that can occur"*. The secrecy and lack of information intended to prevent misuse of IDTs also affected and interfered with the objectives of licensees (P3, P4, P9, P10, P13). While museums employed staff to specifically accompany and support visitors during their interaction with IDTs, they were refused access to documentation and further material. Out of options, the staff resolved to build their own bodies of knowledge through experimentation: *"I think something that would be really helpful is if we had access to either the database of questions that they were asked or some kind of guide to like categories of questions they were asked. Anything like that. Because right now it's very exploratory" (P10).* We observed diverging and sometimes conflicting knowledge on IDT functionality and available IDT responses between different sites or even different staff at the same site. Additionally, publicly sharing their own material was not always possible. P13 reported being explicitly forbidden from utilizing their own recordings of interactions with IDTs at their site: *"I'd wanna be able to put clips on TV spots. I'd wanna be able to put clips on social media. Right now the licensing won't allow any of that"*. Similarly, during our own visits to these educational sites, we were frequently dissuaded from creating any type of recording of any exhibited IDT for any purpose.

Partially motivated by recent pandemic-related restrictions and mandated temporary closures, some IDTs have gradually been made more broadly available online, which complicated and conflicted with controlling and limiting control of access (P10, P17, P22, P23). This systemic change led to reducing or relinquishing previous access restrictions. *"Especially now that we are simultaneously working on online access. At some point, we can no longer control which institutions have access" (P22).*

#### 3.2.2.4 Discussion and Implications

**Prepare for revising IDT implementations.** Unlike other ECAs, IDTs can not rely on AI-generation to add answers or output modalities after the opportunity to record more authentic content has passed. Considering advances in technology, static implementations are expected to be unable to stand the test of time and keep up with user wants or needs. Consequently, the creation and recording of IDTs requires particular foresight and provisioning in order to continue to appear plausibly realistic [Tra16; Glo21; Mar+22], durable [Odo+09], interpretable [JA18], and use curiosity-eliciting technology [Kam18] by future standards. Providing and evaluating IDTs early, i.e., before the deaths of their respective witnesses, can allow for identifying issues while they can still be remedied, e.g., missing additional functional responses. When recording contemporary witnesses for IDTs it can also be advantageous to capture as many modalities in high data quality as possible, even if displaying these qualities is not yet practical [Art+14; Tra+15b]. Displaying the respectively highest output quality available for each individual channel could maximize the perceived naturalness, engagement, empathy, and credibility [Par+22]. These recordings also need to be accompanied by arrangements to ensure their reliable, safe, and secure long-term storage to facilitate the creation of new IDT iterations in the future. Additionally, the input processing system will require continuous revisions to reflect the language changes and developments over time [MCW15], e.g., if the ASR fails to recognize non-binary gender forms [SU21].

**Create and provide varied versions to cater to individual needs.** User needs and preferences are diverse and, at times, incompatible, which means that a single, universal IDT design is not ideal or even accessible to every user. Instead of artificially limiting users by providing a single, fixed interaction method, IDTs would benefit from offering alternative combinations of input and output modalities and levels for users to choose from [RA19]. This also warrants awareness and focused consideration of all prospective users' needs in order to avoid inadvertently excluding entire user groups. Adaptive implementations could individually augment user experiences without detracting from the experiences of others. While inspired by and intended to simulate traditional survivor talks, IDTs can unintentionally reproduce their real-world limitations. Consequently, the potential of the digital and virtual nature of IDTs is frequently not fully utilized with regard to improving accessibility as well as the interaction overall without compromising authenticity.

Incorporating automatic language recognition and translation of the input [VAP19] as well as optional subtitles or closed captions could afford users to converse with IDTs in their preferred language. Simultaneous automatic gesture recognition [Lee06] could enable IDTs to afford and react to non-verbal conversational input, e.g., a user bowing as a greeting or waving goodbye. Furthermore, automatic sign language recognition [Er-+17] would empower non-speaking users to address IDTs directly and naturally.

At the same time, digital methods add their own new limitations, which need to be taken into account. The quality and accuracy of machine-learning-based processing of user input, for example, can be inconsistent for different types of users due to biases of the underlying training data [Koe+20], which neglect entire user groups [CLL23] and need to be alleviated.

Additionally, IDTs can be required to engage users with diverse levels of prior interest and knowledge. Instead of relying on external supplies of necessary information to users, IDTs benefit from being able to self-dependently provide the appropriate context, e.g., a recording of the witness retelling their life story.

**Reduce systemic barriers.** The development and application of additional location-independent versions of IDTs are necessary to detach IDTs from the sparse set of sites providing access. While we encountered similar reluctance regarding public availability as a previous study [Mar+22], our study shows a gradual shift towards allowing additional less restricted, browser-based access to select IDTs. Although embedding IDTs in easily accessible websites improves their reach, it is not an all-encompassing solution.

On the one hand, we encountered multiple instances of insufficient Internet access during our study. Regions without sufficiently stable connections to the Internet necessitate IDTs that can operate offline, like locally instantiated copies. These would require a decentralized approach as well as processes to regularly re-establish consistency across the instances.

On the other hand, the hardware available to the recipient, such as the user's own consumer-grade devices, might afford only a small subset of the offered input and output modalities and thereby provide a diminished experience in comparison to specialized setups, i.e., not as lifelike [Brü18; BBG19].

Consequently, dedicated portable hardware configurations that can be brought or shipped to schools would represent a practicable third approach to further improve reach and availability of IDTs while offering a similar immersiveness and user experience as stationary installations.

Finding ways to make the production and operation of IDTs more affordable and less labor-intensive would reduce the frequently prohibitive associated financial barriers. Worries about jeopardizing a highly expensive investment lead to an increased need for control and inhibit openness about information. Operational or contractual dependencies from the entity which produced and hosts an IDT can result in encompassing restrictions, which impact users, museum staff, and researchers alike. The restrictions can result in a lack of critical information, the prohibition of recording and sharing interactions with the IDT, or the inability to independently adjust UI parameters, like the RGB value of a button.

Additional IDTs of more contemporary witnesses by further interested groups would not only increase the number and diversity of preserved stories and conversations but also lead to more diverse designs, accelerate research, and foster collaboration.

### 3.2.3 Unanticipated Emotional Functions

While not directly pertinent to the focus of this thesis, we included this theme and our corresponding analysis as a basis for discussion and to encourage future work. Although emotional connections between users and IDTs were desired, our participants reported emotional functions and effects that they did not anticipate. We detail instances of unilateral, para-social relationships with the displayed Holocaust survivors and the use case of IDTs as a tool for processing grief.

#### 3.2.3.1 Para-social Relationships

Our participants described incidents of users developing one-sided, para-social relationships [RM87] with IDTs as well as with the witnesses represented (P4, P8, P16, P21, P26). Interactions with IDTs were perceived as shared experiences, which resulted in unidirectional familiarity and trust toward the human witnesses, who were entirely unaware of the user and their encounter. P21 exemplified how prior interactions with the IDT influenced the attitudes and behaviors of users who met the real contemporary witness afterward: *“All these people were giving him these hugs and [greeting him]. He’d never seen them in his life. It was like one person after the next went, ‘Hi!’, ‘Hi!’, ‘Hi!’. And he was going, ‘Who are you?’ ”.* The unequal footing of these encounters developed from the emotions and intimacy users perceive during interactions with IDTs, which they also transfer onto the real human beings. The participants also reported instances of people attempting to bond with IDTs by sharing their own trauma. However, the IDTs were not equipped to acknowledge and validate feelings or appreciate and reciprocate their trust. Exchanges with the corresponding human witnesses were not rewarding either, as they were unaware of what the users shared with their digital replica. IDTs were unable to replicate the bidirectional, lasting relationships emerging from conversations with human witnesses, like Holocaust survivor P26: *“I went to Poland for the first time and I met a group of young people and two of them still keep visiting me at home. And we are close friends, although they are my grandchildren’s age. So there are links formed in this way. What has the [IDT] to do with this?”*

These para-social relationships did not build on misleading users into believing that they were interacting with real human beings. As museum staffers, P4 and P8 had extensive experience and substantiated knowledge about the technical nature and inner workings of IDTs. Both were acutely aware that there was no relational counterpart that shared their feelings of intimacy and familiarity. Nevertheless, they treated IDTs as part of their friend groups and also astonished the corresponding human witnesses by greeting them disproportionately cordially at their first meeting.

#### 3.2.3.2 Processing Grief

Two of our participants, P17 and P21, are children of contemporary witnesses of whom an IDT was created and who have passed in the meantime. This put them in the rare position of being able to have posthumous conversations with the digital twins of their late parents. Both participants expressed appreciation for this opportunity and have continued to interact with the respective IDTs to reminisce and commemorate their deceased family members as part of their grieving processes. Despite being closely familiar with the views and experiences of their parents during their lifetimes, the participants gained new information and insights from talking with the IDTs: *“I learned many things about my own identity that I did not know before” (P21).* Similarly, P17 recounted their emotional experience of assuming to know every available answer and discovering unexpected replies. At the opening of an exhibition, they prompted the IDT of their late mother to share a story about her child P17:

> *"I thought it was gonna be a fun little response. Well, in 1987 I had cancer and I almost died. And she starts talking about that and within 10 seconds [I'm] crying like crazy on camera in front of tons of people. And so that's probably one of the most significant things that's happened, particularly since my mom passed away. [...] It was so incredibly lifelike, it brought back those memories I had forgotten. [...] It brought back pain, but it brought back the joy of having a mother and a father who, you know, got me through a really tough time in my life. [...] I still cry when I hear it 'cause I've [demonstrated] it for other people and I try to be numb to the words, which are not easy, but yeah, it obviously affects me. Yeah. But they're good tears. They're good tears."*

We also encountered an instance of a widow finding solace in visiting the IDT of her late husband, which was detailed by museum staffer P8: *"It must have been an unbelievable mixture of a much more heightened sense of [feeling lucky that we could still interact with him] plus a mixture of the sadness and the loss. So I said, 'I'm sorry if that was difficult'. And she said, 'No, this is why I came.' "*

Our participants acknowledged that predicting the needs or preferences of future bereaved can be difficult or uncomfortable (P17, P19, P20). Since people grieve differently and attitudes toward IDTs vary, the aforementioned experiences were not universal. Whereas P17 expressed regret over not having a similar opportunity to interact with their late father, their sibling has never been interested in engaging with the IDT of their mother, because *"it freaked [the sibling] out"*. Additionally, the perceived value and allure of IDTs of family members or friends could shift after their passing. Due to these uncertainties, IDTs were described as a precautionary comforting measure that might turn out not to be needed by some and missed by others: *"People keep voice messages from their parents after they pass away, they keep recordings, you keep videos, it's the same thing. You don't have to watch it, you don't have to listen to it. But it's there if you want to. And it's there for you to show other people in your family and other generations who those people were" (P20).*

#### 3.2.3.3 Discussion and Implications

**Desire for bidirectional connections.** Conversations with IDTs aim to reproduce social interactions, which are also reciprocal: Users desire to not only request and receive information but also share their own experiences with their conversational partners. In concurrence with the CASA, this also includes users who are well aware of the technical, non-human nature of IDTs. Simulated bidirectionality would include providing appropriate reactions to users recounting potentially emotional parts of their lives. Instead of considering and reacting to such user input as out-of-scope, future IDTs could detect the emotional state of the user and display a pre-recorded, fittingly validating response.

An alternative approach to satisfying users' desire for emotional affirmation and connection would be providing an opportunity to relate to other human users. IDTs could serve as a waypoint for visitors to asynchronously connect with other users by allowing them to record and share their own experiences at the exhibition site [Liu+18].

**Consider future impact on bereaved.** Posthumous interactions with IDTs by people who were close with the corresponding contemporary witness and emotionally affected by their death present an intricate use case. It also furthers our understanding of how technology developed for other purposes is used for coping with grief [Mas+10]. Previously, prospective users' personal familiarity or history with a witness was mostly neglected due to their low proportion.

Future implementations can benefit from considering both the subjective advantages for and the potential tolls on surviving family members and friends. On the one hand, we showed that IDTs can be a valuable asset for the bereaved [Pat+23; Xyg+23; KB08]. They can provide a last opportunity for users to receive answers to questions about themselves that they were unable to ask during the lifetime of their relative or friend, which supports the transformation and continuation of their relationship [MB11]. IDTs can also support re-experiencing or sharing memories and associated emotions. On the other hand, the desire to interact with an IDT while grieving was neither universal nor constant. IDTs in public settings, for example, could inadvertently expose grieving people to a lifelike digital reproduction of the deceased, needlessly reminding them of their loss.

## 3.3 Limitations

While our work found several new insights on IDTs, it has some limitations:

- We focused our study on IDTs of Holocaust survivors. IDTs of witnesses of other historical events might have different or additional requirements and qualities. Future research could explore the characteristics of IDTs of other historically marginalized communities or of people who experienced joyful historical events, e.g., winning a world championship.
- We recruited people involved with the creation or use of IDTs of Holocaust survivors. Consequently, our sample may be skewed toward people with favorable views of IDTs or heightened interests in arguing for their broader production and application. Future studies could focus on more critical perspectives, e.g., by gathering data from Holocaust survivors who deliberately chose not to have an IDT of theirs created.
- Our sample does not represent and reflect every person involved with IDTs. Frequently, Holocaust survivors of whom an IDT exists could no longer be reached for interviews due to their declining health or death [Mar+22]. We also faced difficulties getting in contact with bereaved who continued to interact with the IDT of their late relative or friend, as this is a very narrow group of private individuals.
- Interviewee behavior might have also been adversely affected by their awareness of our familiarity with the topic. Despite our repeated inquiries, participants might have omitted details or incidents they considered mundane or known.
- While experts can serve as surrogates for a larger group [BLM09], data from interviews are not immune to biases and inconsistencies due to their recall-based nature [Ala18]. Consequently, subsequent targeted user studies on IDTs are warranted to further investigate and advance the issues raised in our research.

## 3.4 Extended List of Requirements

Our study and corresponding analysis confirmed as well as expanded our prior list of requirements for IDTs that emerged from our literature research (see Section 2.3). Our findings particularly reinforced the requirements R2: Conversational Input, R4: Accessibility, R5: Revisability, and R6: Self-sufficiency, which we revised to reflect our new insights. Additionally, we identified two new requirements, R7: Adaptability and R8: Self-contained. We highlighted the additions to the list of requirements below in blue. As before, the list of requirements represents ideals that implementations can and should strive for but are unlikely to realize fully.

- **R1: Authenticity:** The content and reactions of IDTs need to be authentic, i.e., recordings of actual statements by a contemporary witness instead of synthetic or AI-generated reactions. Additional arrangements may be required to allow the authenticity of an IDT to be verifiable.
- **R2: Conversational Input:** IDTs need to afford natural conversational user input, which can be both verbal as well as non-verbal. This includes robust and timely recognition and processing so that each input can be matched with an appropriate output.
- **R3: Lifelike Output:** The digital representation of the contemporary witness should be displayed in such a way that it appears plausibly realistic and natural to users.
- **R4: Accessibility:** To facilitate comprehensive and equal accessibility, IDTs need to support diverse input and output methods (e.g., different languages) and social settings to meet user needs.
- **R5: Revisability:** To continuously extend and improve usability, IDTs need to afford the ability to monitor, review, and adjust their system behavior. Additionally, IDTs need to provide for compatibility with and integration of future technologies.
- **R6: Self-sufficiency:** Users should not have to rely on intermediaries or other types of external support to have meaningful and comprehensible interactions with IDTs.

- **R7: Adaptability:** In order to meet the different and often conflicting needs or preferences, the conditions for interacting with an IDT (as afforded by R4: Accessibility) should be dynamically adaptable or customizable to the user.
- **R8: Self-contained:** Partially or fully self-contained versions of IDTs are necessary to avoid systemic dependencies, such as location dependence or Internet dependence.

With the revised list of requirements, we revisited the compliance of previous IDT design models. Neither the model by [Tra+15b] nor the model by [MCW17] address or contain provisions for Adaptability or Self-contained. We amended our prior overview (see Table 2.1) accordingly, which can be seen in Table 3.2.

**Table 3.2:** Overview of which of the revised requirements are met by which previous model. ✓ indicates fulfilled requirements, ✗ indicates unfulfilled requirements, and ✓/✗ indicates requirements that are addressed, but only partially fulfilled.

| **Requirement** | **[Tra+15b]** | **[MCW17]** |
|---|---|---|
| R1: Authenticity | ✓/✗ | ✓/✗ |
| R2: Conversational Input | ✓ | ✓ |
| R3: Lifelike Output | ✓ | ✓/✗ |
| R4: Accessibility | ✗ | ✗ |
| R5: Revisability | ✓/✗ | ✓/✗ |
| R6: Self-sufficiency | ✗ | ✗ |
| R7: Adaptability | ✗ | ✗ |
| R8: Self-contained | ✗ | ✗ |

## 3.5 Summary

In this chapter, we extended the previous partial answers to **SQ-1:** "What are the requirements to digitally preserve interactive conversations with contemporary witnesses?" and **SQ-2:** "What are the associated practices or shortcomings of current IDT implementations?". We conducted an empirical exploration to further our understanding of the current state of IDTs and identify improvement opportunities. We gathered diverse perspectives and experiences from 26 experts on how IDTs of Holocaust survivors are applied and used in practice. We found that users experience several difficulties during interactions with IDTs and that accompaniment by human facilitators is a pragmatic but incomplete solution. Our results show how IDTs can be a more accessible alternative to conversations with human contemporary witnesses and, in turn, how access to IDTs is limited in numerous ways. We discovered the desire to use IDTs for emotional processes beyond their original design. With the help of this study, we gained valuable insights, which allowed us to expand and complement the list of requirements for IDTs.

# Chapter 4

# Conceptual Model of IDTs

*"[The Analytical Engine] might act upon other things besides number, were objects found whose mutual fundamental relations could be expressed by those of the abstract science of operations, and which should be also susceptible of adaptations to the action of the operating notation and mechanism of the engine. Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent."*

— Ada Lovelace, *1843* [14]

## Contents



In this chapter, we present a conceptual model of IDTs and describe which properties and components are responsible for implementing the requirements laid out above. The model outlines the choices, provisions, and decisions that need to be considered when implementing IDTs. We describe, detail, and discuss our model in Section 4.1. The chapter is summarized in Section 4.2.

## 4.1 Characteristics of the Conceptual Model

In our conceptual model, we combine the components

- C1: Input Device
- C2: Preprocessor
- C3: Matching System
- C4: Database of Recordings
- C5: Archive
- C6: Display

into a system that offers and emphasizes the overarching qualities

- Q1: Modularity
- Q2: Instantiability.

[14] [Lov43], p. 694

Due to its abstraction, the model can not only be applied to all-inclusive, universal designs [IS03], but also offers the necessary flexibility for adaptive, customizable designs, which can further support engagement and sense of agency among users [Sun+12]. Such adaptive designs are necessary when a single universal solution is not feasible, e.g., due to competing or conflicting requirements or preferences. Instead of relying on a single, one-size-fits-all approach, adaptive designs incorporate multiple distinct versions that offer different implementations and variations of the components.

In the following, we individually examine the components of the conceptual model, along with the overarching properties that apply to the entire model. Additionally, we explain which aspects of the model are responsible for implementing which of the previously identified requirements (see Table 4.1). The listing and description of the requirements can be found in Section 3.4.

**Table 4.1:** Overview of which requirement is implemented by which component or property of our conceptual model.

| Requirement | Implemented by |
|---|---|
| R1: Authenticity | C4: Database of Recordings, C5: Archive |
| R2: Conversational Input | C1: Input Device, C2: Preprocessor |
| R3: Lifelike Output | C4: Database of Recordings, C6: Display |
| R4: Accessibility | C1: Input Device, C2: Preprocessor, C4: Database of Recordings, C6: Display |
| R5: Revisability | C2: Preprocessor, C3: Matching System, C5: Archive, Q1: Modularity |
| R6: Self-sufficiency | C3: Matching System, C4: Database of Recordings |
| R7: Adaptability | Q1: Modularity |
| R8: Self-contained | Q2: Instantiability |

### 4.1.1 Components

In this subsection, we discuss the individual components of the conceptual model. Figure 4.1 shows the interaction between the components and the associated data flows. The components C1: Input Device and C6: Display represent the User Interface, while C2: Preprocessor, C3: Matching System, and C4: Database of Recordings constitute the corresponding System tasked with the actual processing of the user input. C5: Archive is used to initialize and populate C4: Database of Recordings. In implementations of the model, however, the boundaries between the components may be less strict.

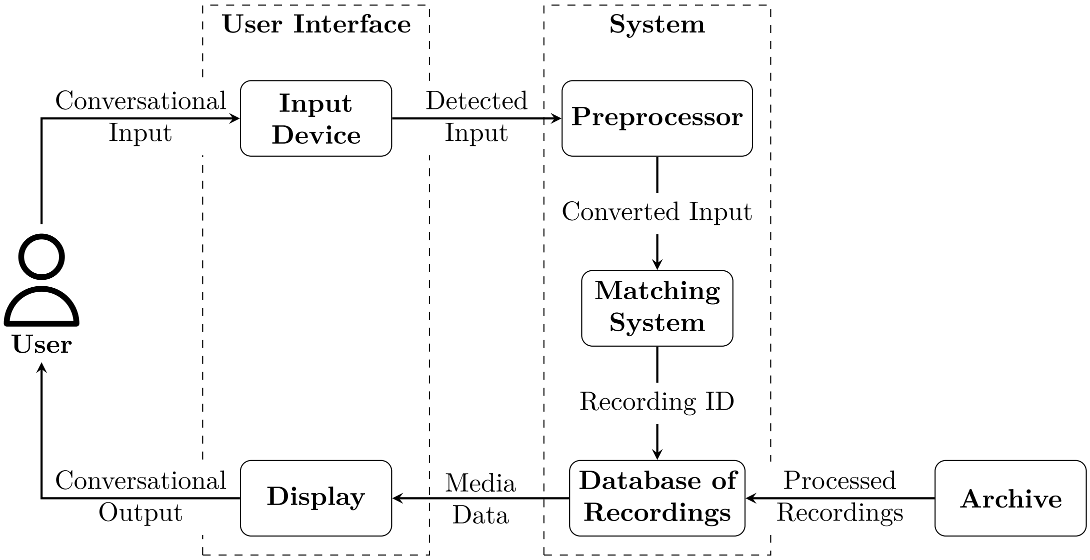



**Figure 4.1:** Conceptual model of IDT components and their respective functions.

#### 4.1.1.1 C1: Input Device

This component combines sensors, processing units, and accompanying software to recognize and record conversational input by users. This includes segmenting user input by detecting the beginning and end of each input in an unobtrusive, timely, and robust manner. We use $I_I$ to denote the set of options afforded by the Input Device, e.g., the different input methods, modalities, or social settings. Example implementations may include, but are not limited to, microphones, physical or digital buttons, keyboards, gaze trackers, or gesture trackers. Coupled with C2: Preprocessor, this component implements the R2: Conversational Input requirement and contributes to the fulfillment of R4: Accessibility.

#### 4.1.1.2 C2: Preprocessor

The Preprocessor receives the detected data and prepares it for processing by C3: Matching System by interpreting and converting the digital signals or measured values into one or multiple formats that ensure readability by machines and humans. Human readability and overall transparency are essential for observing and monitoring the behavior of this component and consequently compliance with R5: Revisability. We use $I_P$ to denote the types of input that the Preprocessor is able to accept and interpret. Examples of preprocessing are automatic recognition and conversion of speech or gestures into text or automatic language detection. Together with C1: Input Device, this component is responsible for implementing the R2: Conversational Input requirement and contributing to fulfilling R4: Accessibility.

#### 4.1.1.3 C3: Matching System

Based on the preprocessed input, the Matching System determines which entry in C4: Database of Recordings is considered the most serviceable to the user. This can be a specific answer to a given question, but also context, guidance, or support for users. The ability to continuously monitor, modify, and improve the component is necessary for accurate and usable matches. Consequently, the Matching system assists in fulfilling the requirements R5: Revisability and R6: Self-sufficiency.

#### 4.1.1.4 C4: Database of Recordings

The Database of Recordings contains the entire set of IDT outputs to be displayed. These are exclusively media files of pre-recorded actions, reactions, or responses by the original human witness, which is key to fulfilling the requirement R1: Authenticity. The ability of the IDT to conversationally handle boundary conditions or software or hardware limitations results from the availability of dedicated functional recordings within the Database of Recordings. These can include a supportive introductory testimony, an inviting idle loop, and informative fallback replies. An example of a complete interaction session that uses such functional recordings can be seen in Figure 4.2. Meaningful and comprehensible functional recordings enable users to interact with the IDT fully independently, which is necessary to fulfill the R6: Self-sufficiency requirement.

As the database provides all the recordings that are to represent the contemporary witness, multiple versions of each recording are available. By providing diverse modalities, resolutions, or levels of detail, the experience can flexibly be tailored to different users or output devices. This supports both R4: Accessibility and R3: Lifelike Output. We use $O_{Rec}$ to denote the aforementioned set of output options afforded by C4: Database of Recordings.

#### 4.1.1.5 C5: Archive

The Archive provides safe, secure, and long-term storage for the original and unmodified recorded data. This serves two purposes: On the one hand, it allows inspection of the complete original context of each recording and, thereby, verification of the authenticity of the IDT output. On the other hand, the archive is the source for all current and future contents of C4: Database of Recordings, as it stores all recorded data, including, but not limited to, modalities or resolutions not currently in use. This enables revisions, updates, or additions to C4: Database of Recordings by processing and preparing the archived data for effective and efficient streaming and output. The Archive therefore helps to fulfill the requirements R1: Authenticity and R5: Revisability.

#### 4.1.1.6 C6: Display

This component comprises output devices and software that present the media data provided by C4: Database of Recordings to the user in a plausibly realistic, conversational, and natural way. This also includes handling the transitions between the individual media sequences. We use $O_D$ to denote

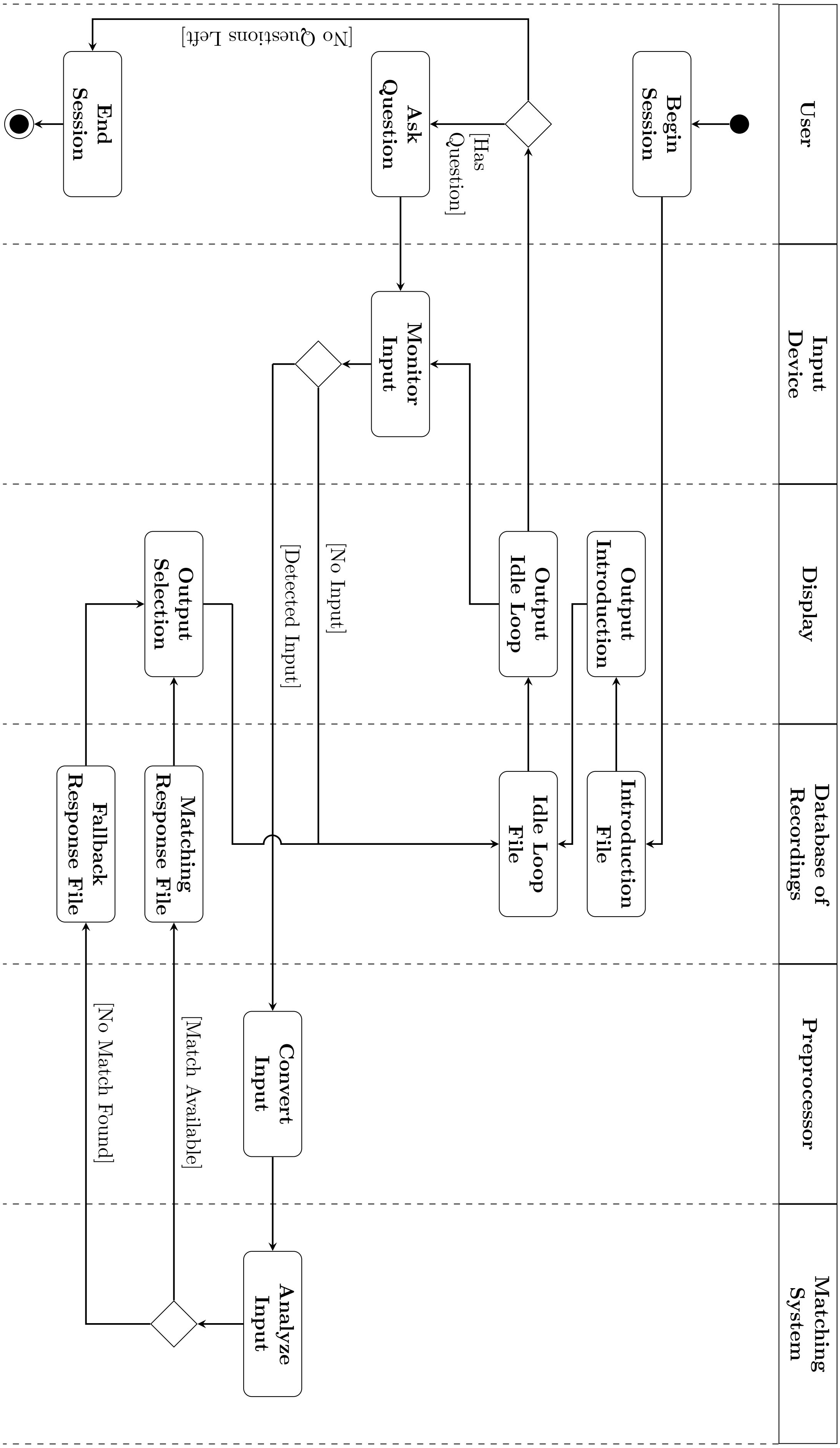



**Figure 4.2:** Partitioned UML 2.5.1 Activity Diagram of an example interaction with an IDT. It shows how functional recordings such as Introduction, Idle Loop, and Fallback Responses can be incorporated.

the set of options afforded by the Display, e.g., the different output methods, modalities, or social settings. Example implementations may include, but are not limited to, speakers, headphones, 2D screens, volumetric displays, solo interactions, or group settings. Together with C4: Database of Recordings, this component is responsible for implementing the R3: Lifelike Output requirement and contributing to fulfilling R4: Accessibility.

### 4.1.2 Q1: Modularity

The modular design of IDTs allows individual components to be modified and seamlessly exchanged or replaced without restricting the functionality of other components or the entire system. Standardized interfaces and data types arrange for compatibility between the components.

The modular structure comes into play when preparing for future technical developments and their integration into an IDT. This may involve, for example, the continuous improvement of C3: Matching System or the expansion of the set of utilized types of C6: Display to include previously unused modalities available in C5: Archive. Modularity therefore contributes to the fulfillment of R5: Revisability.

Additionally, addressing the conflicting individual preferences or capabilities of users may necessitate multiple simultaneous but different implementations of subsets of the model. The input and output options available to a user are determined by the affordances of both the system as well as the respective user interface (see Figure 4.3). The provision of a specific input format requires that both a compatible C1: Input Device is available and that C2: Preprocessor can interpret said format. Likewise, provisioning a specific output format requires that it is both contained in C4: Database of Recordings and that a compatible output device is available. Employing a modular design and providing diverse C1: Input Devices and C6: Displays allows users to dynamically adapt their experience with the IDT to their own needs, which realizes the requirement R7: Adaptability.

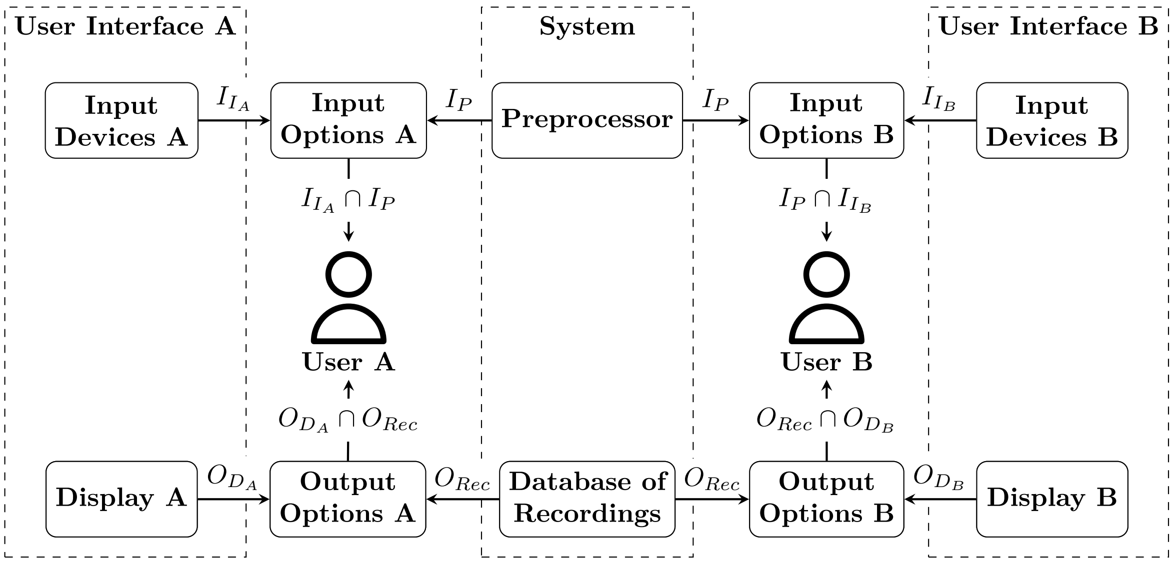



**Figure 4.3:** A modular IDT configuration with two distinct user interfaces. $I_{I_A}$, $I_{I_B}$, $I_P$ describe the input options afforded by Input Device A, Input Device B, and Preprocessor, respectively. $O_{D_A}$, $O_{D_B}$, $O_{Rec}$ describe the respective output options afforded by Display A, Display B, and Database of Recordings. By employing different implementations of the user interface, User A and User B can be offered complementing input and output options while using the same centralized processing system.

### 4.1.3 Q2: Instantiability

The ability to create local instances enables IDTs to bypass systemic dependencies, such as a mandatory internet connection during the interaction. These instances can be local copies of individual components or of the entire system. A local instance that encompasses all components required during the interaction with an IDT (i.e., all but C5: Archive), represents a fully self-contained system (see Figure 4.4). This allows the user and system to be in the same place, eliminating the need for exchanging data with remote components. Consequently, IDT instances can be delivered to users in regions with insufficient digital infrastructure instead of users having to travel long distances to interact with an IDT at a specialized location. Instantiability fulfills the requirement for IDTs to be R8: Self-contained and can help to balance and distribute the load on each individual component. However, to minimize deviations from a centralized main system and avoid inconsistencies, a consistent method or protocol is required according to which the local instances are updated or replaced.

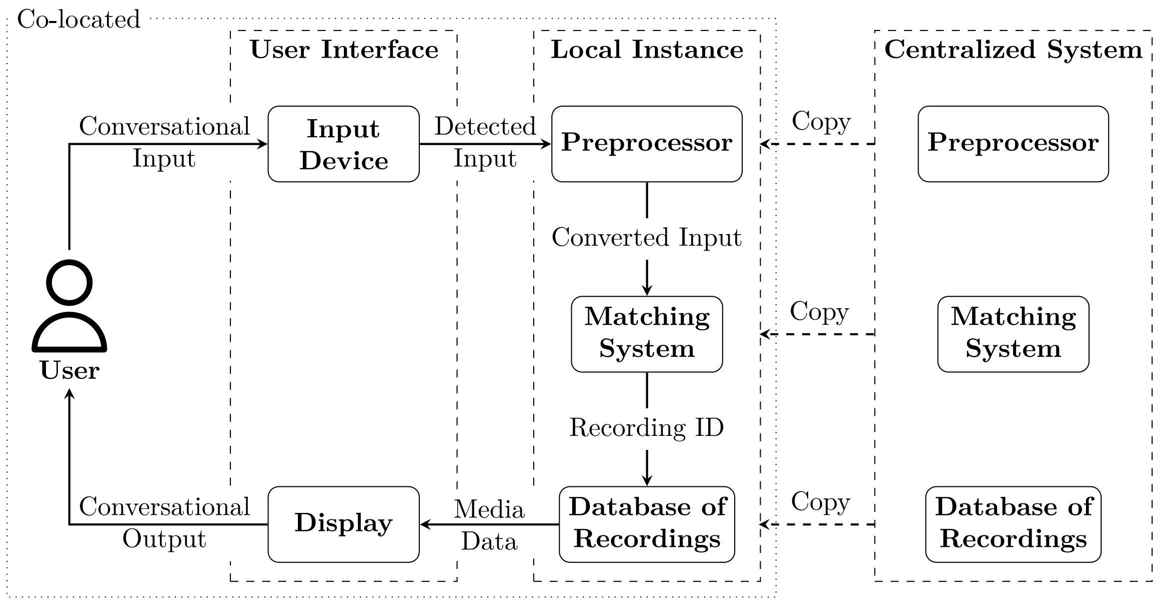



**Figure 4.4:** Creating copies of all components of the processing system (Preprocessor, Matching System, and Database of Recordings) allows users to interact with an IDT without requiring a stable connection to remote components, e.g., of a Centralized System. This setup constitutes a fully self-contained IDT.

Infrastructural circumstances permitting, it can therefore be advantageous to instantiate only a subset of the processing system, i.e., C2: Preprocessor, C3: Matching System, and C4: Database of Recordings. Examples of two different, partially self-contained instances of an IDT can be seen in Figure 4.5. The depicted setups demonstrate how different parts of the processing system can be instantiated locally.

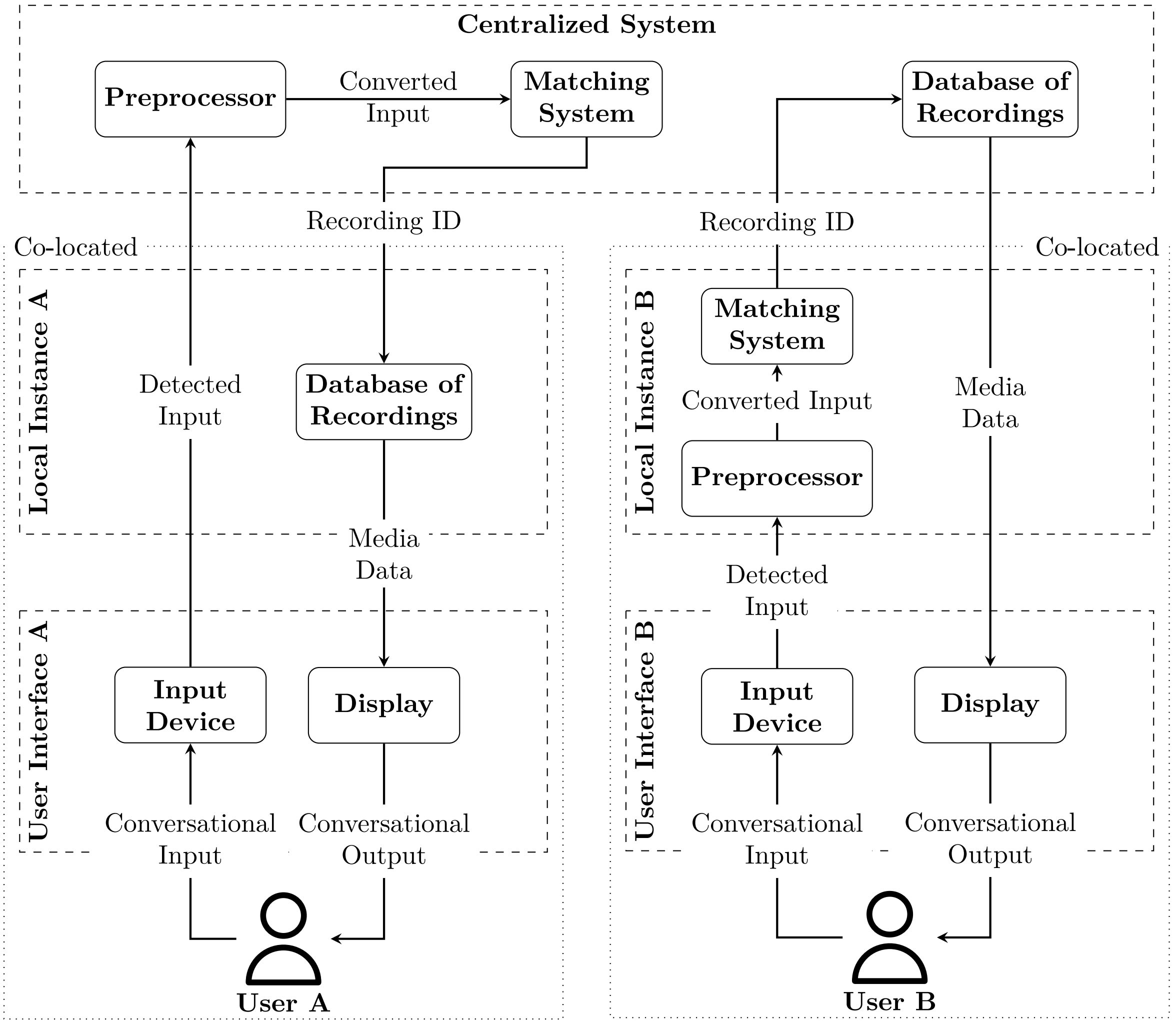



**Figure 4.5:** Two examples of different partially self-contained IDT setups. Setup A (left) uses a local copy of the Database of Recordings. Setup B (right) uses local copies of the Preprocessor and the Matching System.

## 4.2 Summary

In this chapter, we answered **SQ-3:** "How can IDTs be designed to fulfill the identified requirements?". We detailed how an IDT design consisting of C1: Input Device, C2: Preprocessor, C3: Matching System, C4: Database of Recordings, C5: Archive, and C6: Display and building on both Q1: Modularity and Q2: Instantiability can realize all the requirements that we identified during our previous requirements analysis. Our conceptual model of IDTs therefore exceeds the previous models, which we examined in Section 2.4. In our description of the structure of the model, we discussed the individual components, their tasks, the model-wide properties, and explained which aspect satisfies which requirement.

# Chapter 5

# Example Implementation: LediZ

*"While I was touring the camp I encountered three men who had been inmates and by one ruse or another had made their escape. I interviewed them through an interpreter. The visual evidence and the verbal testimony of starvation, cruelty and bestiality were so overpowering as to leave me a bit sick. In one room, where they were piled up twenty or thirty naked men, killed by starvation, George Patton would not even enter. He said he would get sick if he did so. I made the visit deliberately, in order to be in position to give first-hand evidence of these things if ever, in the future, there develops a tendency to charge these allegations merely to 'propaganda.'"*

— Dwight D. Eisenhower, *April 15th, 1945* [15]

## Contents



In this chapter, we present an example implementation of our conceptual model of IDTs and the corresponding design choices. This implementation covers the life stories of and by Holocaust survivors Abba Naor [NZ14] and Eva Umlauf [UO16] and represents the very first IDTs in the German language. These IDTs were developed jointly as part of the interdisciplinary project LediZ.

In the LediZ project, we use a three-pronged implementation of IDTs (see Figure 5.1): A stationary high-fidelity instance with a focus on immersiveness, a transportable instance that aims to retain as much as possible of the stationary instance, and an online instance that can be retrieved via browser, for example. A centralized system provides C2: Preprocessor, C3: Matching System, and C4: Database of Recordings, which also serve as a ground truth for the instantiations. Both the stationary instance and the transportable instance use their own local copy of C4: Database of Recordings, whereas the online instance utilizes all the components provided by the centralized system. This modular approach is facilitated by our use of technical standards as well as standardized interfaces and data formats. C5: Archive stores the unmodified recorded data, whose processed versions populate C4: Database of Recordings within the centralized system.

[15] The Papers of Dwight David Eisenhower, The War Years IV, doc #2418

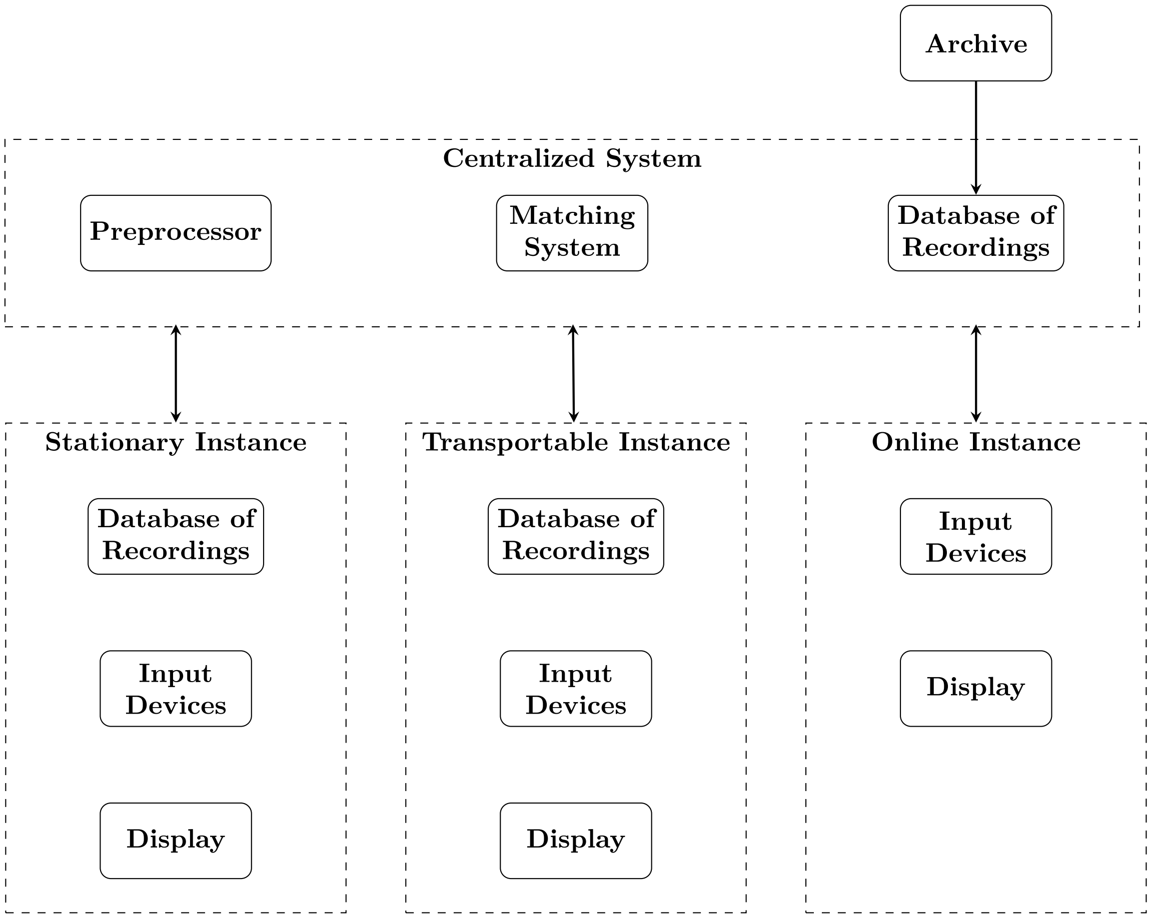


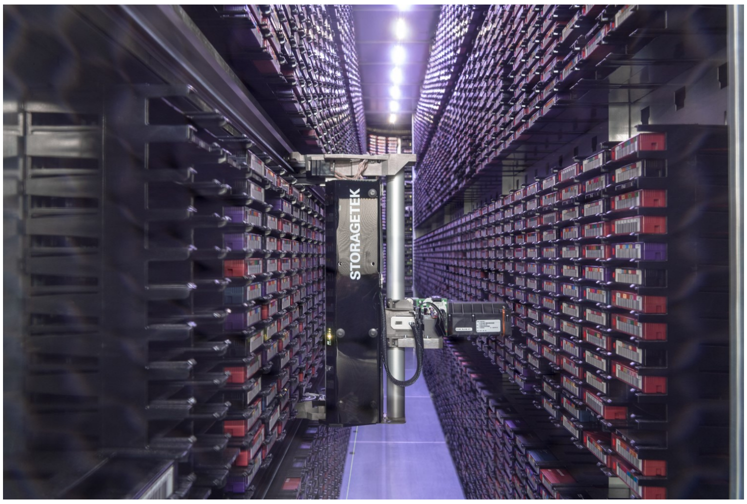

**Figure 5.1:** LediZ simultaneously offers three ways for users to interact with IDTs: A stationary instance, a transportable instance, and an online instance. No instance is fully self-contained, as each instance requires an Internet connection to the centralized system.

## 5.1 Archive

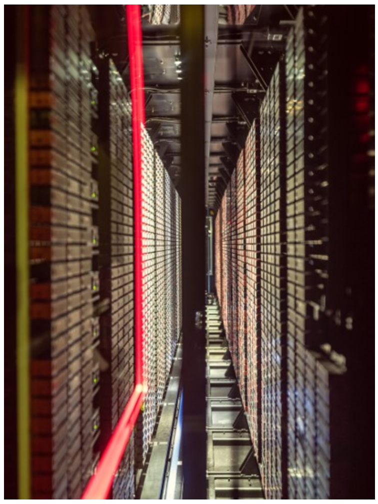
The original recorded data of both contemporary witnesses exceeded 32 TB in total. This stemmed mostly from the high-resolution stereoscopic visual modality, which was simultaneously captured by two synchronized RED Epic-M Dragons using DRAGON 6K S35 sensors[16]. The raw footage from each camera used a 6K HD resolution (5568 × 3132 px), a frame rate of 50 Hz, and Redcode 12:1 compression.

The entire set of raw data is stored in the tape libraries of the LRZ (see Figure 5.2), which are operated using the IBM Tivoli Storage Manager. This archive implements a variety of safety precautions[17] to ensure the safe, secure, and long-term storage of the recordings. The data are stored in encrypted form and with protected access for editing and modification. The assured service level agreements include uptime and availability of more than 95% throughout the year. The raw recordings are stored redundantly, including an additional copy of each tape at the LRZ as well as a further mirror off-site at the Max Planck Computing and Data Facility (MPCDF)[18]. The data are also specifically flagged for indefinite storage and will be retained even during system migrations, as the recordings will be copied to each subsequent archive system.

[16]https://docs.red.com/955-0181_v7.4/DSMC2_DRAGONX_7_4/Content/A_TechSpecs/Specs_DSMC2_DRAGONX_6K.htm
[17]https://doku.lrz.de/benutzungsrichtlinien-11475999.html
[18]https://www.mpcdf.mpg.de/




**(a)** Picture by Andreas Heddergott.

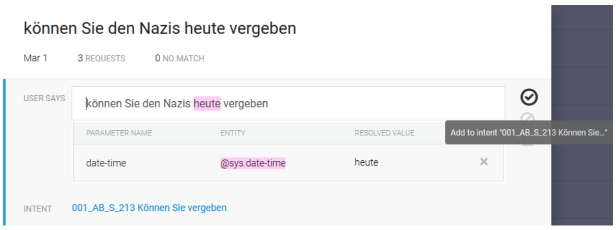


**(b)** Picture by Robert Brembeck.

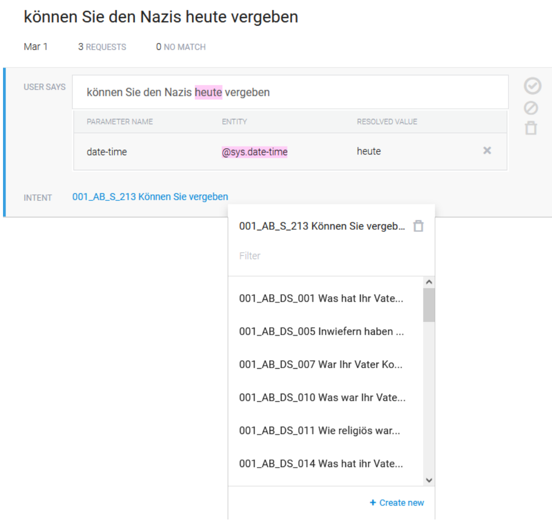
**Figure 5.2:** Automated tape library at the LRZ, in which the raw recordings that were created as part of the LediZ project are stored.

## 5.2 Centralized System

The LediZ IDTs of Abba Naor and Eva Umlauf use the current release of Google Dialogflow, which provides both speech-to-text as well as intent matching, to implement both C2: Preprocessor and C3: Matching System at once. While this proprietary cloud-based service aims to ensure a fast processing time as well as a monthly uptime of more than 99.9%[19], its inner workings and processes are only traceable and explainable to a limited extent. However, it readily facilitates supervised learning by allowing staff members to monitor and label input data as well as the concluding decisions of C3: Matching System (see Figure 5.3). It is not possible to create one's own local or even offline instance, which, however, also avoids possible version conflicts when training C3: Matching System.




**(a)** Confirming that the intent returned by Dialogflow matches the user input.




**(b)** Reassigning user input to different intent.

**Figure 5.3:** Manually labeling user input data in Dialogflow to train the matching system.

The central database is hosted at the LRZ and, at the time of writing, managed with OpenCast 9.3[20]. It contains 948 recordings for the IDT of Abba Naor and 993 recordings for the IDT of Eva Umlauf, which are computed from the higher-resolution raw data in C5: Archive. Each recording exists as a Full HD 2D (1920 × 1080 px) and side-by-side stereoscopic 3D version (3840 × 1080 px),

[19] https://cloud.google.com/dialogflow/sla
[20] https://docs.opencast.org/

with a respective frame rate of 25 Hz and 44.1 kHz mono audio. The database also contains occasional variants of recordings, including different lengths of introductory testimonies and diverse sets of fallback replies targeting different reasons why no other matching answer can be provided. However, due to technical limitations, these can not be retrieved by users, since the processing system currently does not support returning alternative versions for similar or identical inputs.

The Wowza streaming engine 4.8.5[21], at time of writing, provides the media data via the HTTP live streaming protocol. Additionally, a custom-made Resolver assists in mapping the recording identifiers returned by C3: Matching System to the uniform resource locators used by the streaming service, which eases compatibility between these components.

## 5.3 Stationary Instance

For the first prong of our three-pronged implementation, we use the Powerwall[22] (see also Figure 5.4) of the Centre for Virtual Reality and Visualisation (V2C), at the LRZ. This Powerwall is an Infinity Wall-like screen [Cze+97] and provides a projection area of up to 6 $m \times 3.15$ $m$, though we only utilized the proportion necessary to display the IDTs in life-size. Similar to an Infinity Wall [Cze+97], the Powerwall displays stereoscopic images by superimposing and circularly polarizing the respective images for the viewer's left and right eye. The superimposed images are then filtered using passive 3D glasses with the appropriate polarization filters, i.e., clockwise and counterclockwise. Consequently, the Powerwall represents a partially immersive, VR system [Muh15]. The two projectors, which deliver the respective images for the right and left eye, are each capable of achieving a brightness of up to 21 000 ANSI Lumen and a resolution of 4096 × 2160 px. The audio output is provided by a 7.1 surround sound system.

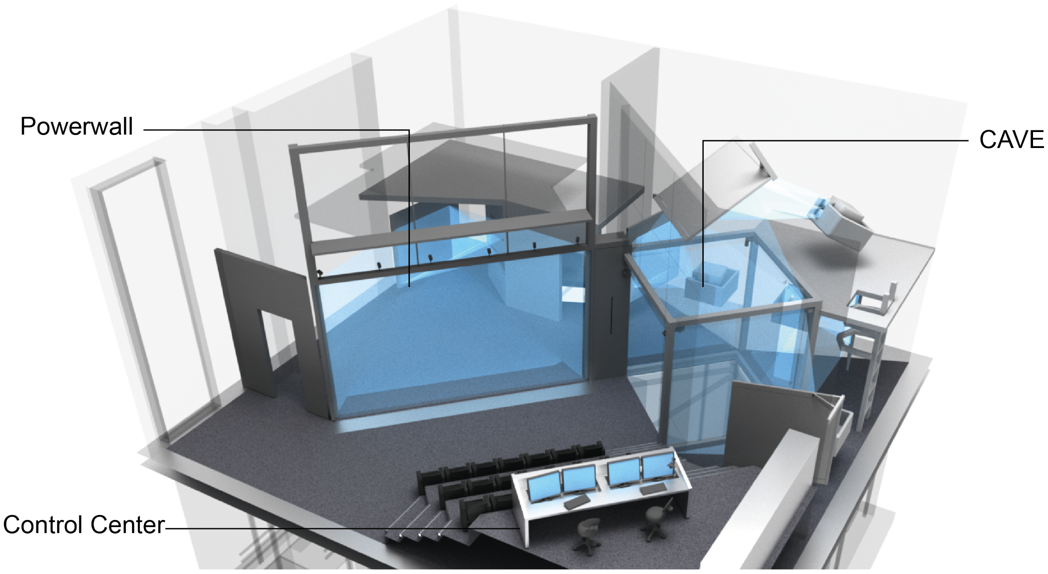
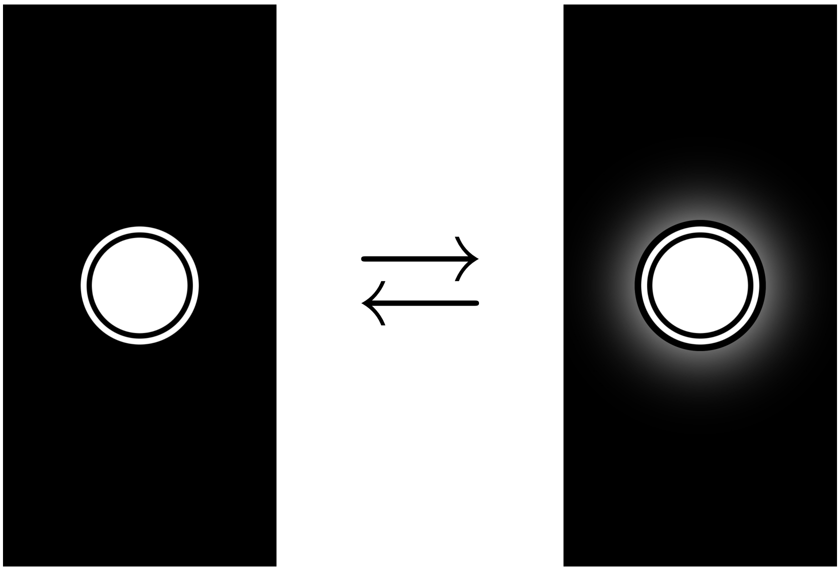



**Figure 5.4:** The Powerwall as installed at the Centre for Virtual Reality and Visualisation (V2C) of the LRZ. This image was previously published in [May+23], ©2023 IEEE. (Also pictured, but not relevant for this thesis, are the Cave Automatic Virtual Environment (CAVE) and the V2C control center.)

As a fixed installation, the overall setup is stationary and not transportable. The V2C itself offers seating for up to 21 concurrent visitors. The available C6: Display devices afford multiple different modes of output, including audio-visual stereoscopic 3D, audio-visual monoscopic 2D, and audio-only.

The stationary instance relies on the centralized system, specifically C2: Preprocessor and C3: Matching System, to analyze the user input. However, this instance uses its own local copy of C4: Database of Recordings in MP4 format (ISO/IEC 14496-14:2020) [Int20], even though the centralized database and streaming server are located nearby. This is because the Internet bandwidth proved insufficient to transmit the 3D videos efficiently. As they are connected via shared networks, updates and changes to the centralized database can also be immediately applied to this local copy.

[21] https://www.wowza.com/docs
[22] https://doku.lrz.de/Powerwall-10615075.html?showLanguage=en_GB



(a) Button in waiting state, which signals that the recording process is not active.

(b) Button pressed, which activates the recording of spoken input.

**Figure 5.5:** Digital button of the mobile input device used to trigger the recording of spoken input.

We use a dedicated smartphone as a mobile C1: Input Device that can be passed around among users. A plain and straightforward graphical user interface (GUI) (see Figure 5.5) allows users to trigger the recording of voice input. The interaction itself can be configured and controlled with the help of a companion application. It implements, among other functions, the selection of the IDT of a specific contemporary witness (see Figure 5.6), changes to the output modality, or the targeted playback of the introductory testimony. However, the configuration of the IDTs is not made directly accessible to users and is instead controlled and managed by trained staff.

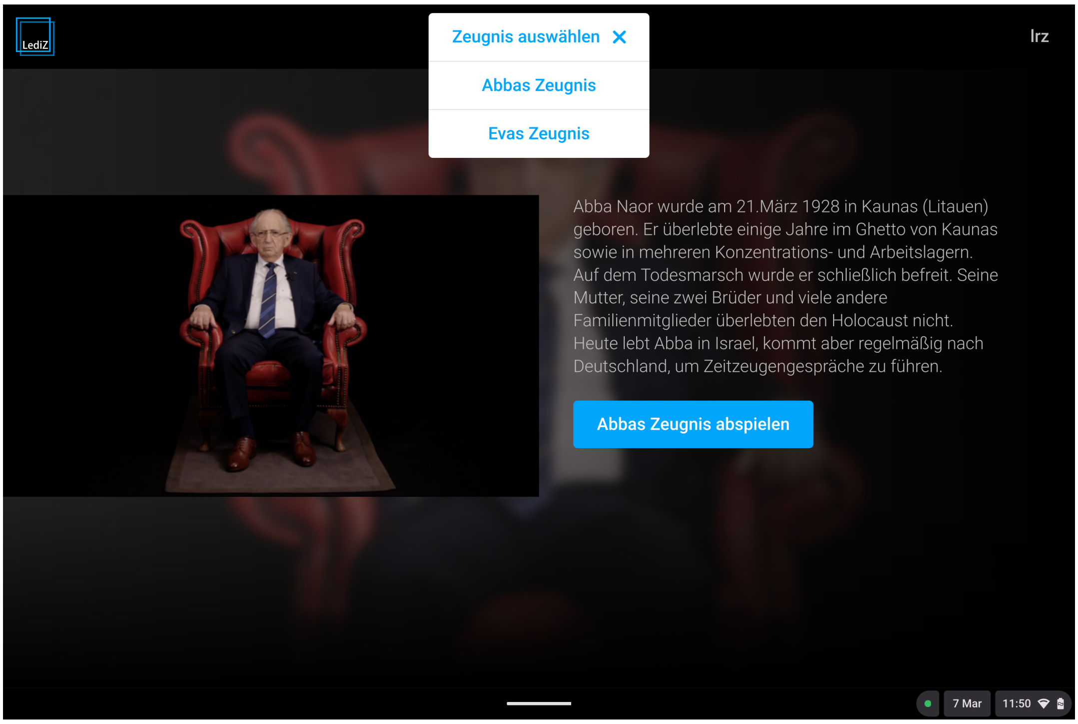



**Figure 5.6:** An accompanying application on a mobile device, e.g., a tablet, can be used to configure and adapt the interaction to the needs and interests of the user. This figure shows the drop-down menu for selecting the IDT ("Zeugnis auswählen", transl. "Select testimony").

## 5.4 Transportable Instance

To also provide users who are not in the vicinity of the LRZ with an immersive experience with IDTs, we created an instance that is transportable and, at the same time, mirrors the aforementioned stationary instance as closely as possible. To achieve this, we use a roll-up projection screen with a width of 266.7 cm and a height of 150.1 cm, which results in a screen diagonal of 304.8 cm and an aspect ratio of 16:9. However, the screen does not support passive stereo vision, as this would require a special coating of the canvas, which is associated with additional restrictions on storage as well as transportation. Consequently, we rely on active stereo vision instead. We realize this with the help of a portable projector[23] that alternates between displaying the image for the left eye and the image for the right eye in Full HD (1920 × 1080 px) and at up to 4000 ANSI Lumen, respectively. Active shutter 3D glasses, whose shutters are synchronized with the switching of the images by the projector, separate the images for the viewer's left and right eyes. Our stock of these shutter glasses allows for up to 50 concurrent viewers. Additionally, the transportable instance utilizes two stereo speakers to provide the audio output of the IDT. Like the stationary instance, the transportable instance consequently affords multiple different modes of output, including audio-visual stereoscopic 3D, audio-visual monoscopic 2D, and audio-only. An example application of the transportable instance can be seen in Figure 5.7.

**Figure 5.7:** Operation of the transportable instance at a remote location. It allows students to interact with the stereoscopic 3D IDT of Abba Naor without having to travel to the LRZ. Picture by Ernst Hüttl.

The transportable instance also relies on C2: Preprocessor and C3: Matching System of the centralized system to analyze user input, as these cannot be instantiated locally in our implementation. Nevertheless, to minimize dependence on the Internet and available bandwidth, the transportable instance uses a laptop and a local copy of C4: Database of Recordings in MP4 format. Updates and changes to the centralized database can be applied to this local copy as soon as a stable Internet connection can be (re-)established.

The C1: Input Devices likewise mirror those of the stationary instance: A dedicated smartphone serves as a mobile device for voice user input. A companion application allows staff members or educators to configure the IDT, such as changing the output modality or selecting the IDT of another contemporary witness.

## 5.5 Online Instance

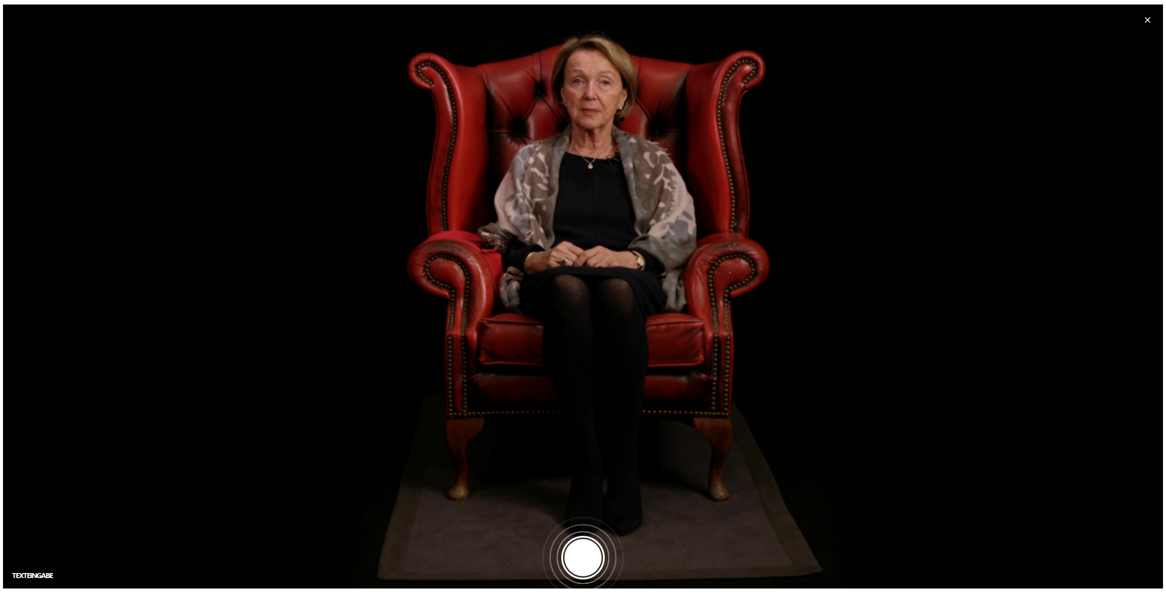
The third prong of our three-pronged implementation is a browser-based online instance for remote users. This allows us to further extend the reach of our IDTs beyond that of the transportable, physical instance. The online instance also requires less effort to scale with the number of concurrent users.

[23] https://www.sharpnecdisplays.eu/p/datasheet/en/datasheet/t/projectors/business-projectors/rp/m403h.xhtml

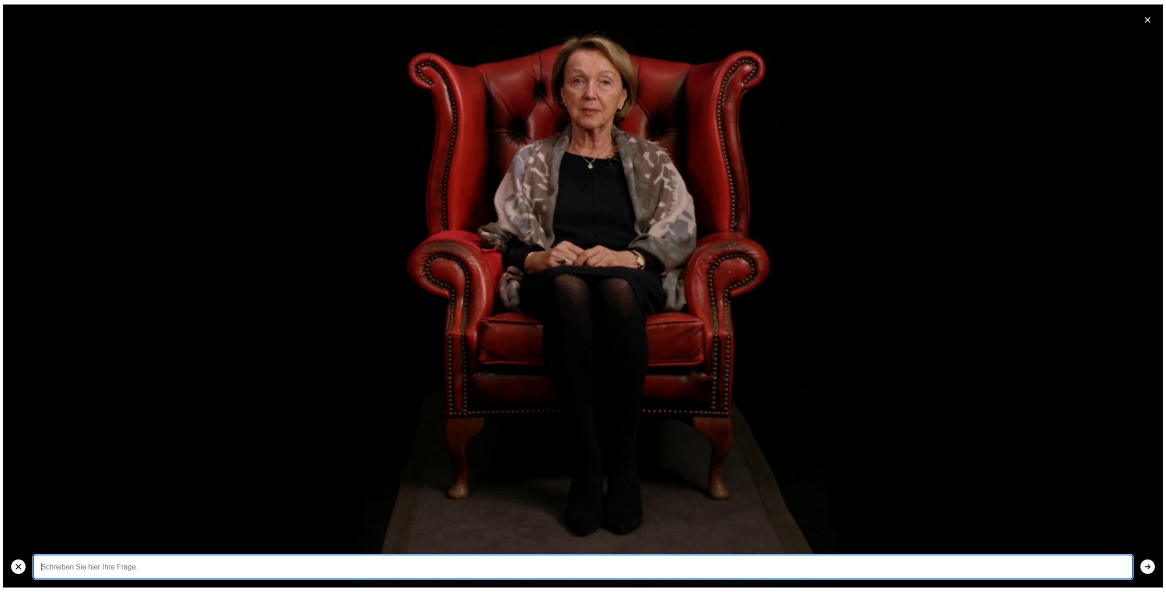
Since only the user's own input and output devices are used, no additional devices or maintenance personnel are needed during the interaction. However, this entails a trade-off with the warranted properties of the interactions with the IDTs, such as their immersiveness. For example, the user's devices may not be capable of visually displaying the IDT in life-size or a particularly immersive manner.

The online instance utilizes the entire centralized system for pre-processing the user input, assigning a matching response, and retrieving the corresponding media. While the Wowza streaming engine of the centralized C4: Database of Recordings is equipped to provide the IDT in stereoscopic 3D, at the time of writing this thesis, the online instance only makes use of audio-visual 2D. This is due to the aforementioned bandwidth problems and the unlikelihood of sufficient output devices on the user side.

The online instance affords users additional flexibility in terms of input modalities. In addition to spoken audio input, it also allows users to provide their conversational input via typed text input (see Figure 5.8). Furthermore, the online instance offers users more freedom regarding the setting of their interactions with the IDT, e.g., on their own without observers and at home in a familiar, comfortable environment.




**(a)** While pressing the digital button, audio input detected by the user's microphone is recorded and subsequently sent to the centralized system for processing.




**(b)** Alternatively, users can also enter their questions in a text field using their keyboard.

**Figure 5.8:** Interactions with the online instance of the IDT of Eva Umlauf.

The modular structure of the overall implementation is particularly important for the online instance. To reach as many users as possible, the online instance needs to be compatible with as many different browsers, operating systems, and input or output devices as possible. This is achieved by the fact that the components can be easily exchanged and that established protocols are used, e.g., standardized file formats and HTTP live streaming.

## 5.6 Fulfillment of Requirements

Our implementation fulfills the following requirements:

- R1: Authenticity by using exclusively pre-recorded responses and reactions and abstaining from AI-generated content. Storing the original, raw recordings and the context of each reaction in a safe, secure, and long-term archive allows for verifying the authenticity of the content of the IDT.
- R2: Conversational Input by supporting both spoken as well as text-based natural conversational user input. By combining the pre-processing and matching steps in a joint cloud-based service, the IDT can reliably and timely process the user input.
- R3: Lifelike Output by combining audio-visual output modalities, including stereoscopic 3D, realistic life-size presentation, and brief crossfades to mask transitions between displayed recordings.
- R4: Accessibility partially by affording multiple input and output modalities as well as social settings, such as alone or in a group. However, at the time of writing, input and output support German language only and lack further accessibility features.
- R5: Revisability partially due to the modular configuration and the retention of the raw recordings in the archive. The overall revisability is bounded by the concrete qualities of the raw recordings as well as the use of a proprietary service for processing the user input, which hinders monitoring and reviewing processing steps. Additionally, the proprietary service limits the amount of accepted training data[24]. However, the use of a proprietary service also exemplifies one of the trade-offs of the implementation, as creating and managing a new NLU from the ground up was not feasible.
- R6: Self-sufficiency partially by incorporating dedicated functional responses and recordings to allow users to interact with the IDT without requiring additional assistance. However, several configurations and settings are not accessible without staff involvement.
- R7: Adaptability by offering multiple different instances as well as diverse options for input, output, social setting, and access, each of which can be adapted or customized to meet user's needs or preferences.
- R8: Self-contained partially, since two of the three instances of our implementation are partially self-contained, reducing their reliance on a centralized system's availability and reachability. Additionally, our transportable instance provides a novel method of supplying users with IDTs that are both immersive as well as not bound to a specific location. However, the remaining dependence on an Internet connection means that areas without any Internet can not be served.

Consequently, LediZ fulfills or partially fulfills each of the previously identified requirements (see Table 5.1). However, this depends on the entirety of the instances, as the individual instances do not fulfill all requirements equally. As we outlined in Section 2.3, any implementation completely fulfilling all requirements is unlikely, since trade-offs may be necessary.

[24] https://cloud.google.com/dialogflow/quotas

**Table 5.1:** Overview of which requirements are fulfilled by the LediZ implementation and comparison with previous models. ✓ indicates fulfilled requirements, ✗ indicates unfulfilled requirements, and ✓/✗ indicates requirements that are addressed, but only partially fulfilled.

| Requirement | LediZ | [Tra+15b] | [MCW17] |
|---|---|---|---|
| Authenticity | ✓ | ✓/✗ | ✓/✗ |
| Conversational Input | ✓ | ✓ | ✓ |
| Lifelike Output | ✓ | ✓ | ✓/✗ |
| Accessibility | ✓/✗ | ✗ | ✗ |
| Revisability | ✓/✗ | ✓/✗ | ✓/✗ |
| Self-sufficiency | ✓/✗ | ✗ | ✗ |
| Adaptability | ✓ | ✗ | ✗ |
| Self-contained | ✓/✗ | ✗ | ✗ |

## 5.7 Summary

In this chapter, we exemplified the IDT implementation of the project LediZ, which follows a three-pronged approach, each with its own C1: Input Device and C6: Display: A stationary high-fidelity instance with a focus on immersiveness, a transportable instance that aims to retain as much of the stationary instance as possible, and an online instance that is accessible via a browser, for example. A centralized system, which also serves as a ground truth for local instances, provides C2: Preprocessor, C3: Matching System, and C4: Database of Recordings. Both the stationary instance and the transportable instance use their own local copy of C4: Database of Recordings, while the online instance relies on all the components of the central system. Our use of technical standards, standardized interfaces, and standardized data formats provides the necessary Q1: Modularity and Q2: Instantiability. The C5: Archive at the LRZ stores the unaltered recorded data, the processed versions of which populate C4: Database of Recordings within the centralized system.

As a result of this approach, we implement all the components and overarching properties of the model we presented in Chapter 4 and consequently fulfill or at least partially fulfill each of the previously identified requirements.

# Chapter 6

# Empirical Evaluation: Effects of the Visual Output Modality of IDTs

*"You are my witnesses"*

— Isaiah 43:10 [25]

[Inscription in the Hall of Witness at the United States Holocaust Memorial Museum]

## Contents



In this chapter, we describe our empirical evaluation of the relationships between IDT output modality and user experience, addressing subquestion **SQ-4:** "How does the type of visual output modality affect how users experience the interaction with the implemented IDT?".

We conducted two separate mixed-methods studies in 2021 to investigate how different levels of visual modality (audio-only, audio-visual 2D, audio-visual stereoscopic 3D) affect user perception:

- The first study (Section 6.2) used a 2×2 between-subjects design to compare audio-only with audio-visual 2D in in-person and online settings ($n = 82$).
- The second study (Section 6.3) used a within-subjects design to compare audio-visual 2D with audio-visual stereoscopic 3D ($n = 51$).

[25] English Standard Version Bible. 2001.

Both studies advance the empirical exploration of the vast design space of IDTs. The raw data sets were collected as part of two bachelor theses[26,27] supervised by the author of this thesis.

In Section 6.1, we present our rationale behind the chosen study designs as well as the experimental setup. We detail the composition of the user sample, the design of the study, the measuring method, the procedure, the quantitative analysis, and the qualitative analysis for each study individually in Section 6.2, respectively Section 6.3. We then discuss our findings in Section 6.4 and explain their limitations in Section 6.5. Section 6.6 provides a summary of this chapter.

## 6.1 Method

We split our investigation into two separate studies, each focusing on two modality conditions. This decision was based on having two comparable IDTs at our disposal as well as the challenges in recruitment and overall logistics due to the SARS-CoV-2 pandemic. Covering all three modality conditions in a single study would have entailed significant challenges and concerns with regard to validity: Using a within-subjects design with three modality conditions would have led to undesired learning effects due to repeated interactions with the same digital witness in different conditions as well as participant attrition and exhaustion due to the prolonged duration of the study sessions. Conversely, using a between-subjects design with three modality conditions would have required more participants over a longer study period during a time of already significantly reduced readiness or ability to travel and participate in studies. Additionally, the unpredictable developments of the pandemic situation and the associated restrictions implied the possibility of an abrupt, indefinite discontinuation of each study. To mitigate these risks, we conducted two individual studies and prioritized internal validity within each study over external validity between both studies.

- 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 and the participants' own devices. 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 6.1) of the LediZ project (see Chapter 5). With IDTs of two different witnesses, we reduced the influence of individual characteristics like gender [MWM19], personality, or narrative. Both study designs were individually reviewed and approved by our institution's ethics committee as well as the legal department.

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

**Figure 6.1:** Examples of the visual appearance of IDTs as used in our studies. The white and blue buttons were used when verbally inputting questions.

[26] Patricia Maier: Comparison of auditory virtual humans and audiovisual virtual humans. 2021.
[27] Simona Maiolo: Evaluation of the difference between monoscopic and stereoscopic digital testimonies regarding immersion and emotional reaction. 2021.

All in-person studies took place at the V2C and displayed the digital contemporary witnesses in life-size (see also Section 5.3). At most six of the 21 available seats were used at the same time due to sanitary restrictions. 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 same NLP system to analyze the intent of the voice input and select the most fitting prerecorded response, which was then displayed. We conducted pilot tests to verify the feasibility of our studies.

In-person participation allowed us to monitor users, including their well-being. If a participant displayed signs of mental distress, we were ready to intervene by halting the session, offering support, and carefully attending to their needs to help them calm down. Example steps include determining whether they want to continue, accompanying them to a separate room, listening reflectively, and validating their feelings. After carefully tending to the distressed participant, we would resume the session with the participants who were willing to proceed. However, no such incident occurred in either study.

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.

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

In this 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.

### 6.2.1 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. The allocation of each participant is also shown in Table B.10 of the appendix. 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 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.

### 6.2.2 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 as well as balanced the groups in all conditions to control for potential confounds caused by the individual characteristics of the IDT ($n_{Eva\ Umlauf} = 40$, $n_{Abba\ Naor} = 42$).

### 6.2.3 Measured Variables

**Table 6.1:** Variables measured in the audio-only vs. audio-visual 2D study. Due to the high proportion of non-normally distributed items [TA16], we used Greatest Lower Bound (GLB) [WJ77] to calculate the internal consistency reliability.

| Section | Variable | Items | Reference | Internal Consistency |
|---|---|---|---|---|
| Questionnaire | User Experience | 4 | UEQ-S [SHT17] | 0.50 |
| Questionnaire | Presence | 12 | IPQ [SFR01], WS [WS98] | 0.83 |
| Questionnaire | Pos. Emotions | 10 | PANAS [WCT88] | 0.85 |
| Questionnaire | Neg. Emotions | 10 | PANAS [WCT88] | 0.86 |
| Questionnaire | Accessibility | 4 | WCAG 2.0 [Cal+08] | 0.68 |
| Questionnaire | Preference | 1 | - | - |
| Interview | Reasoning | 7 | - | - |

With our questionnaire, we measured six variables from a total of 41 items (see Table 6.1). 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) [SHT17]. We measured how present and immersed the participants felt with three items (SP3, INV2, INV4) of the igroup presence questionnaire (IPQ) [SFR01] and nine items (5, 6, 8, 9, 21, 23, 25, 26, 32) of the presence questionnaire by Witmer and Singer (WS) [WS98]. The Positive and Negative Affect Schedule (PANAS) [WCT88] was used to evaluate their emotional responses. We based the questions regarding accessibility on the four WCAG 2.0 guidelines [Cal+08]. We also asked users whether they would have preferred the other modality. An itemized overview of our questionnaire can be found in Section B.1. 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. The interview guide used in this study is shown in Table B.5.

### 6.2.4 Study Procedure

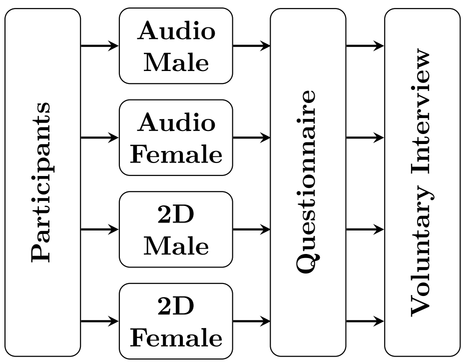



**Figure 6.2:** Study 1 used between-subjects design to compare audio-only and audio-visual 2D IDTs in in-person and online settings.

Both study settings, in-person and online, used the same procedure (see Figure 6.2) 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 twelve 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.

### 6.2.5 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, Analysis 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 [MDK18; Bla+17]. We consequently carried out factorial ANOVA to identify significant ($p < .05$) impacts of modality and setting, as well as interaction effects. Tukey's Honestly Significant Difference (HSD) test was used for post hoc analyses. Table 6.2 summarizes the main findings. The aggregated values and the distribution of preference are shown in Figure 6.3. Our exhaustive quantitative analysis can be found in Section B.3.

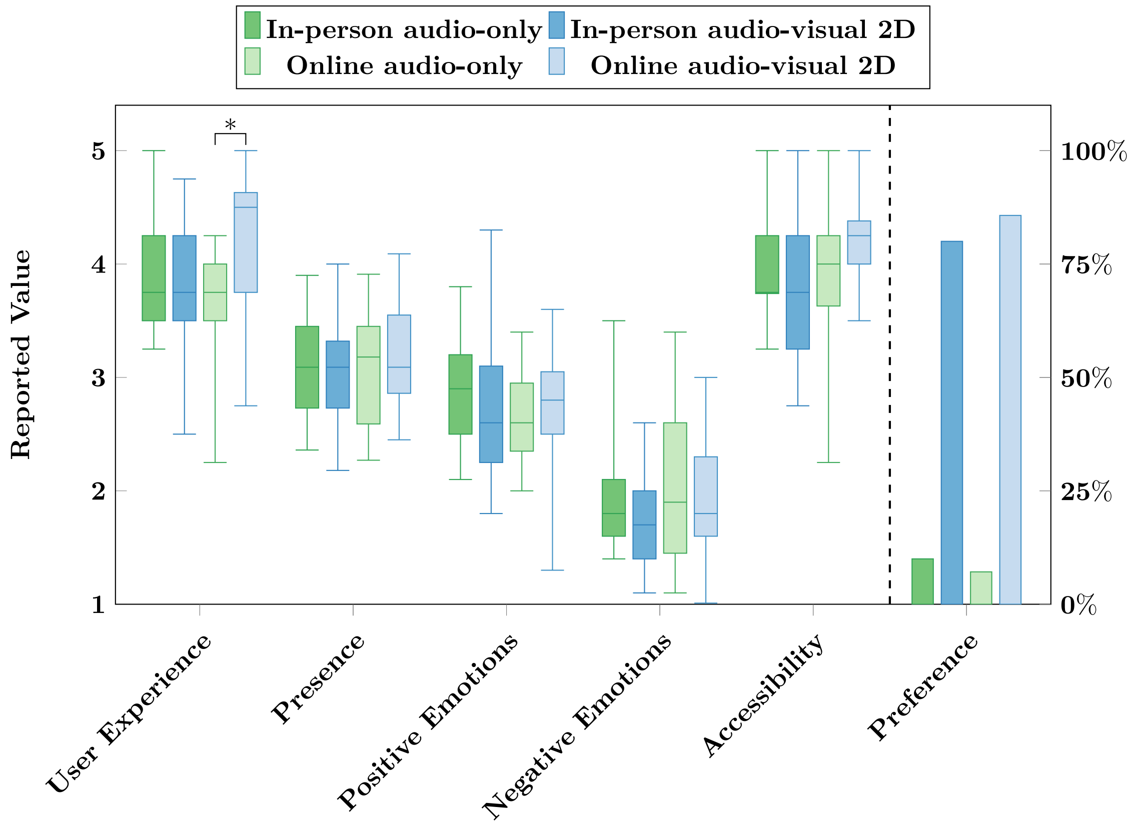



**Figure 6.3:** Aggregated measures of the audio-only vs. audio-visual 2D × in-person vs. online study. * indicates a significant ($p < .05$) pair-wise difference. Separated by a dashed line are the percentages of participants who would have preferred the respectively opposite modality. The totals for preference are less than 100%; the remainder felt indifferent.

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.

**Table 6.2:** 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 Reaction 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) |

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.

### 6.2.6 Qualitative Results

Our reflexive thematic analysis [BC06; BC22] used both the transcribed interviews as well as the free-text explanations on the questionnaires from all four groups. The author of this thesis manually coded the data and constructed 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 [Byr22]. 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 (see also Figure 6.4):

#### 6.2.6.1 Tactful Visuals Promote Engagement and Emotional Connection

Participants across all groups reported enhanced cognitive and affective engagement as well as social presence [SWC76] 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

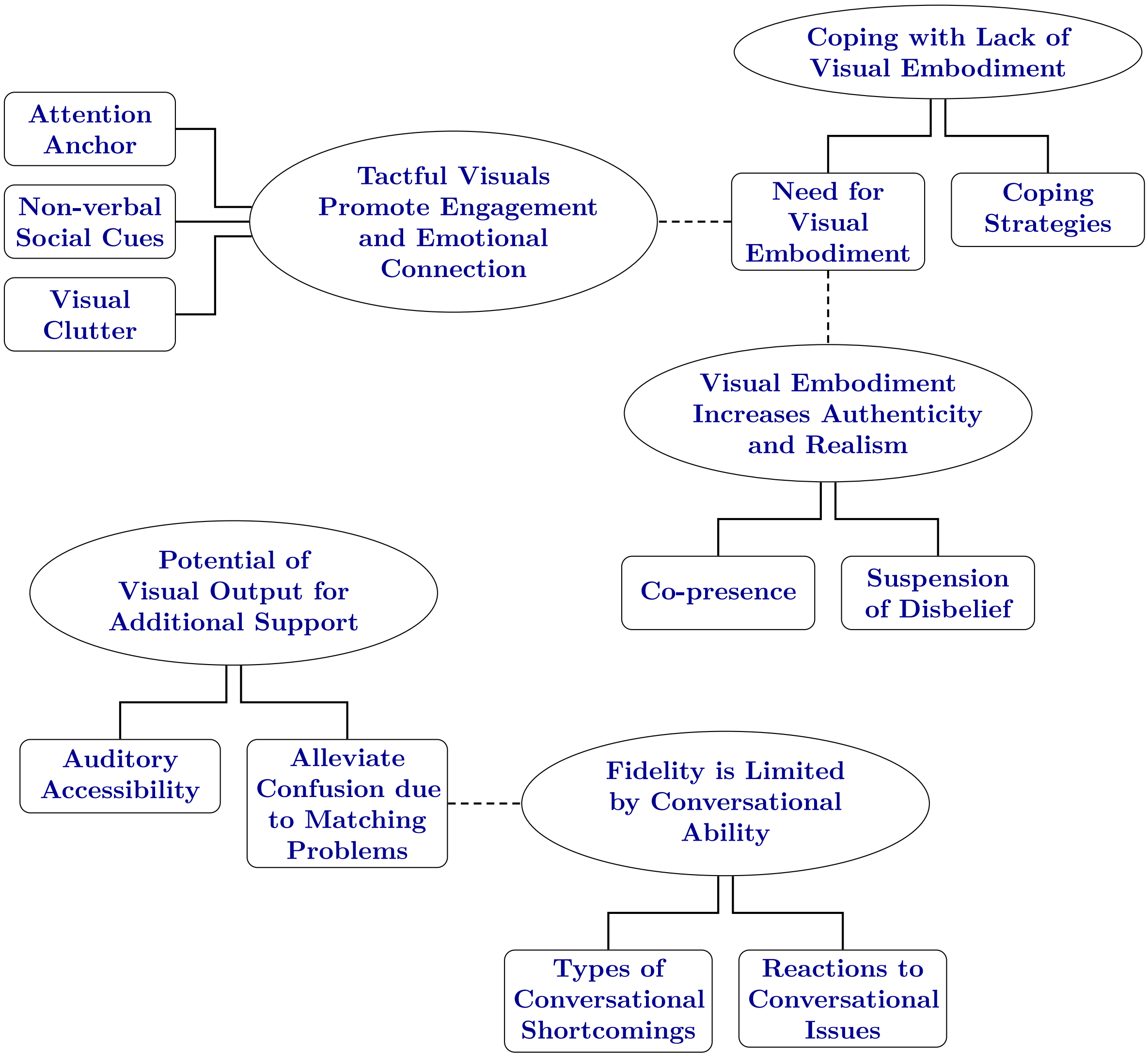



**Figure 6.4:** The themes and sub-themes derived from our analysis of interviews and free-text explanations of Study 1.

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 to talk 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) [DL86], this also demonstrates that the visual output supplies users not only with auxiliary emotional cues but also functional cues like information about turn-taking [Cas00].

**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, 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.

#### 6.2.6.2 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.

#### 6.2.6.3 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 [Gof63] 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).*

#### 6.2.6.4 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 IDT 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:

- 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 IDT 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).*
- 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.

#### 6.2.6.5 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 6.2.6.1 and Section 6.2.6.3, they could also divert attention from the digital witness as well as emphasize the non-human nature of the conversational partners.

## 6.3 Study 2: Audio-visual 2D vs. Audio-visual Stereoscopic 3D

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

### 6.3.1 Participants

We recruited 51 participants through university newsletters and social media channels. We offered them either extra university credit for their studies or 20€.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. The study sequence of each participant is shown in Table C.5.

### 6.3.2 Study Design

We conducted a between-subjects study with *Modality* as 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.

### 6.3.3 Measured Variables

With our questionnaires, we measured five variables from 29 items. The structures and origins of the questionnaires are presented in Table 6.3. 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

**Table 6.3:** Variables measured in the audio-visual 2D vs. audio-visual 3D study. Due to the high proportion of non-normally distributed items [TA16], we used GLB [WJ77] to calculate the internal consistency reliability.

| Section | Variable | Items | Reference | Internal Consistency |
|---|---|---|---|---|
| Questionnaire A | User Experience | 4 | UEQ-S [SHT17] | 0.77 |
| Questionnaire A | Presence | 14 | IEQ [Jen+08] | 0.97 |
| Questionnaire A | Emotions | 6 | AEQ [Pek+11] | 0.89 |
| Questionnaire A | Discomfort | 4 | SSQ [Ken+93] | 0.75 |
| Questionnaire B | Preference | 1 | - | - |
| Interview | Reasoning | 6 | - | - |

on the UEQ-S [SHT17], 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 [Jen+08] 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 (AEQ) [Pek+11], due to its reliability and suitability for immersive virtual environments [Tch+16]. 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) [Ken+93]. After interacting with both modes, Questionnaire B asked users to select their preferred modality. An itemized overview of our questionnaires can be found in Section C.1. 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. The interview guide of this study is shown in Table C.3.

### 6.3.4 Study Procedure

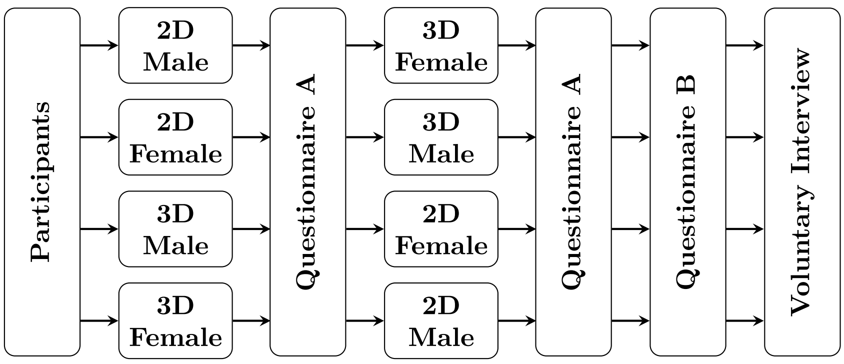



**Figure 6.5:** Study 2 used within-subjects design to compare audio-visual 2D with audio-visual 3D IDTs.

The procedure of this study is presented in Figure 6.5. 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 eleven minutes for the 2D display and twelve 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.

### 6.3.5 Quantitative Results

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 *Emotions* are normally distributed. The aggregated values and the distribution of preference are shown in Figure 6.6. Table 6.4 summarizes our main findings. The complete results of our quantitative analysis can be found in Section C.3.

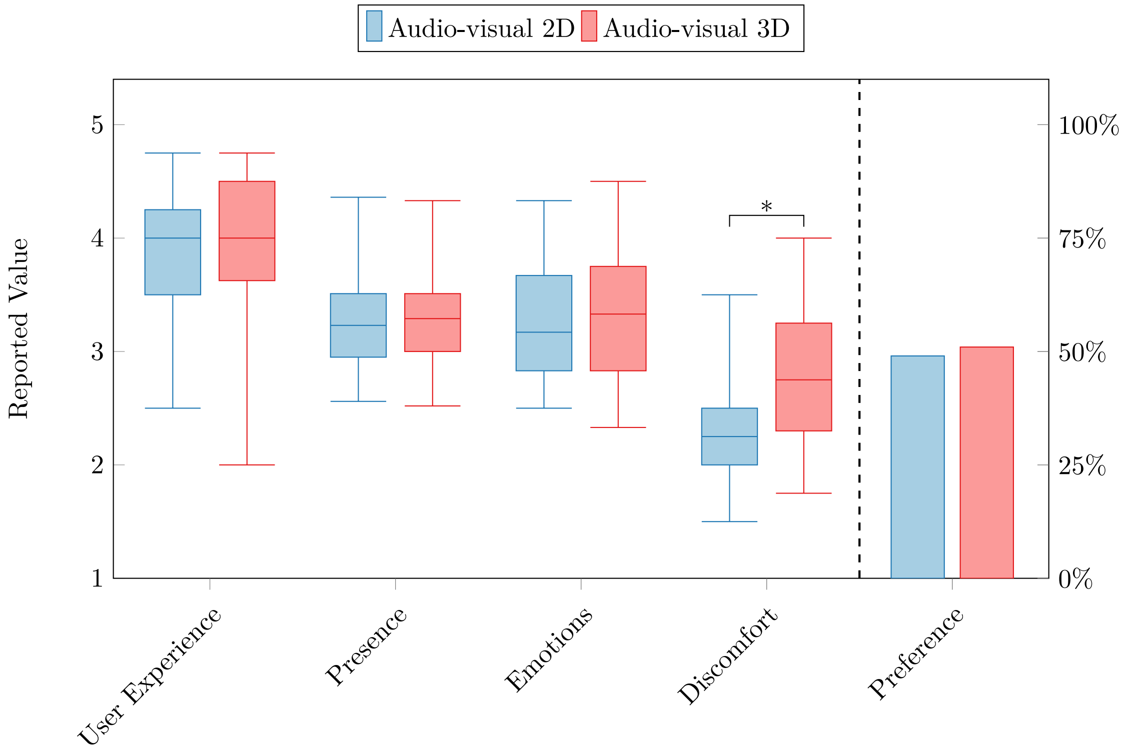



**Figure 6.6:** Aggregated measures of the audio-visual 2D vs. audio-visual 3D study. * indicates a significant ($p < .05$) difference. The total for preference is 100%.

**Table 6.4:** 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) |

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 "just feel like we're in the same room". 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.

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

### 6.3.6 Qualitative Results

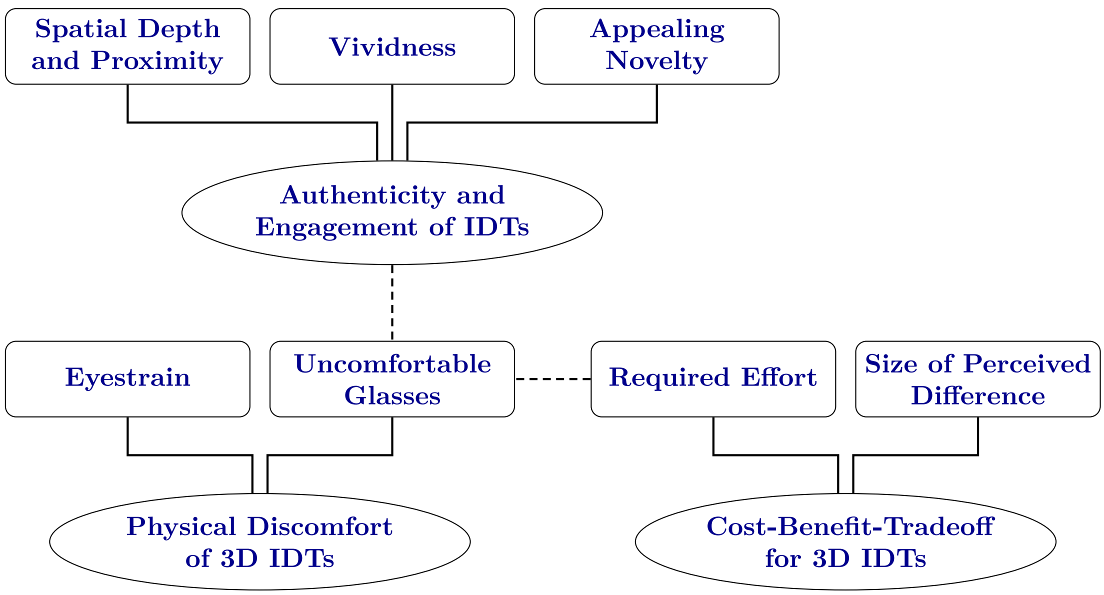



**Figure 6.7:** The themes and sub-themes derived from our analysis of interviews and free-text explanations of Study 2.

We used the same qualitative method as for Study 1: The author of this thesis investigated the transcribed interviews and participants' explanations on the questionnaire using manual inductive coding of primarily semantic meanings and reflexive thematic analysis [BC06; BC22]. The approach was mainly experiential and constructionist [Byr22]. 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 6.7.

#### 6.3.6.1 Authenticity and Engagement of IDTs

The 3D 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 and provided little benefit.

**Vividness:** The 3D 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.

#### 6.3.6.2 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 and engage with the IDT. However, several participants felt annoyed, distracted, and less present. It hindered particularly 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 lead 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.

#### 6.3.6.3 Cost-Benefit-Tradeoff for 3D IDTs

While reflecting on their experiences, the participants also discussed the effort required for the application 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 of acquisition 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 [Glo21], 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 6.3.6.1 and Section 6.3.6.2, 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 6.6). 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).*

## 6.4 Discussion

In this section, we discuss the combined results of both studies. Our main findings are:

- Visual representations of IDTs provide important non-verbal social cues, which improve engagement and focus among users.
- Users face challenges such as uncertainty due to delayed responses in audio-only interactions, highlighting the importance of visual feedback.

- While 3D IDTs can enhance realism and presence, the use of 3D glasses causes physical discomfort and undermines user preference.
- Increased realism in audio-visual IDTs raises user expectations, which can lead to false affordances due to current technological limitations in processing and handling conversations.
- Effective IDTs require consistency in realism across all modalities and must address discomfort and accessibility issues to maximize user experience.
- We recommend capturing testimonies in high-quality multi-modal formats, as it allows for flexibility in future applications. However, practical considerations must balance quality with cost and effort.

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 [NST94] and MRT [DL86]. 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 6.1).

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 [May14]. 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 [Sho+22] 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 [Gav91], and a Habitability Gap [Moo17b]. 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 [Vis+22]. Implementations need to consider these conversational abilities since they represent a major limitation for the realism and immersiveness of IDTs.

Our quantitative analysis found no significant differences in aggregated user experience, immersion, emotiveness, and user preference between 2D and 3D IDT displays. 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 [Mal+15], 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 [Got+18], 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.

Effective and convincing concepts require consistency in realism beyond output modality [NG99; IN00; Moo17b]: 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.

Tactful and unobtrusive visual displays facilitate emotional connections and improve user experience without detracting from the digital witness and their story. However, implementations need to consider the surroundings of the site of operation, as these can influence the effectiveness and limit the available modalities and their levels. Likewise, attention to potential barriers to use, including language and pronunciation, helps to ensure accessibility.

Building on the “maximization” hypothesis [Par+22] 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. 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.

## 6.5 Limitations

While our user studies in Section 6.2 and Section 6.3 found several differences in the effects of display modality, they have some limitations:

- Although the 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 may be 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 [Dah+16]. Similarly, repeating our experiments with IDTs of less traumatic events could investigate whether different, potentially more joyful topics benefit from different presentation and interaction methods.

- 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 possibly distracting influences.

- 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 [Got+18] 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.

## 6.6 Summary

In this chapter, we answered **SQ-4:** "How does the type of visual output modality affect how users experience the interaction with the implemented IDT?"

We investigated the influence of output modality on user perception of IDTs by measuring multiple variables in two distinct user studies, which took place in 2021 during the SARS-CoV-2 pandemic. 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 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 reaffirm 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.

# Chapter 7

# Future Work

*"An allem Unfug, der geschieht, sind nicht nur die schuld, die ihn begehen, sondern auch diejenigen, die ihn nicht verhindern."*
*[All the mischief that happens is the fault not only of those who commit it, but also of those who do not prevent it.]*

— Erich Kästner, *1933* [28]

## Contents



In this chapter, we present directions and approaches for future research on IDTs. Building on our findings and contributions, future work could investigate and develop novel concepts as well as capture and preserve valuable perspectives by other witnesses while it is still possible. Instead of aiming to strictly reproduce the aspects of physical survivor talks, future designs could emphasize utilizing the possibilities of the digital medium. Additionally, IDTs of other groups of witnesses or featuring other languages would provide more representation and accessibility and might uncover additional insights, e.g., an IDT of a contemporary witness using sign language.

Future IDTs could also combine access to multiple witnesses, allowing learners to ask multiple witnesses the same question and compare their responses. Similarly, IDTs could provide answers that were recorded at different stages of the life of the witness. This would provide the opportunity for users to discover whether and how the opinions and attitudes of the witness changed over time.

To better understand the effects of interactions with IDTs on users, further empirical studies are needed, such as comparisons with the effects of conversations with real human witnesses or the consequences of different input and output modalities, e.g., haptic social cues [Fei+19] or olfactory displays [Lan22]. Further research could also discuss the concept of authenticity in the context of IDTs or provide a transparent and uniform delineation of what constitutes misuse of IDTs. Although true future-proof implementations are unlikely [Ale21], research on existing as well as upcoming use cases is essential to advance designs and concepts for preserving the testimonies of current and future contemporary witnesses.

[28] [Käs33], p.103

## 7.1 Further Use Cases

While we focused on the use case of Holocaust education and remembrance during our research, IDTs can also be used in various other contexts, each of which requires and deserves its own in-depth investigation and evaluation. Further areas of application include testimonies by survivors of other traumatic historical events, such as the atomic bombing of Hiroshima [Tak23] or the forced labor camp in Belene [Fou23]. Conversely, IDTs can also be used to interactively experience the narrations of witnesses to non-traumatic or even joyful events, for example, the memories of renowned actors [Sto21], musicians [Sma21], or astronauts (see Figure 7.1), for whom at the time of writing only applications with very limited content and interactivity have been offered.

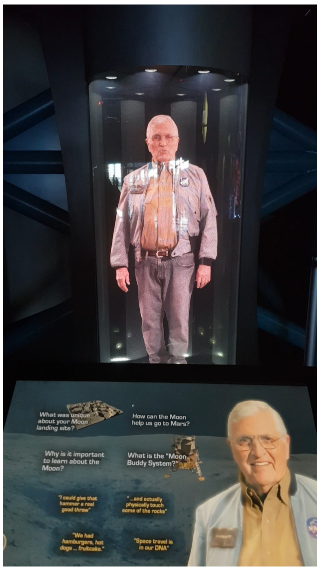



**Figure 7.1:** Virtual reproduction of Apollo 17 astronaut Harrison Schmitt at Kennedy Space Center, Merrit Island, Florida. Visitors can select one of four predefined questions on a touchscreen to receive the corresponding pre-recorded answer.

Another, separate use case is the extension of IDT subjects from specific persons and contemporary witnesses to everyday people. Coping with the death of a loved one is a complex, arduous, and oftentimes enduring process [KWA72]. The loss of a family member or friend has many facets, as this also means, among other things, losing a social connection, shared habits, or a source of knowledge and information. Previous research emphasized the importance of thanatosensitive exploration of how technology and interactive systems can assist individuals in sharing memories or commemorating the deceased, as well as how grieving people adapt existing technology to meet their specific needs [Mas+10]. Systems designed for the bereaved should prioritize the facilitation of storytelling rather than viewing death as the end of the user's relationship with the departed [MB11]. Furthermore, the perceived durability of these interactive systems and the ability to create a personal history through repeated engagement over time have been identified as factors that contribute to the user's overall satisfaction and appreciation of such systems [Odo+09]. Consequently, IDTs of family members or friends can allow users to be proactive and inquisitive listeners who can traverse and explore inter-generational memories on their own terms [JA18].

A further use case is the application of IDTs to help scale services in business or commercial contexts. Especially video calls with narrowly defined topic areas, such as interviews as part of the hiring process [Cha+22] or customer support [Pil11] (see also Figure 7.2) lend themselves to being replaced by IDTs. Further, domain-specific applications exist within the realms of healthcare consultation [JAA21; Moo23] and teaching [Nat22]. Additionally, they can support public relations and company outreach, e.g., with the help of IDTs of otherwise unavailable or hard-to-reach executives [Zei22].


**Figure 7.2:** Web-camera connected information booth, Munich Airport [Pil11]. These interactive kiosks allow visitors to contact customer service staff via video call.

## 7.2 Forthcoming Research Questions

In the following, we discuss eight open research questions on IDTs, which future concepts, frameworks, and designs should consider. These can also advance the understanding, interpretation, and definition of our previously developed requirements (see Section 3.4) as well as further refine our model of IDTs (see Chapter 4).

### 7.2.1 How should IDTs be owned, managed, and controlled?

In many cases, the producers or providers of IDTs retain ownership and near-absolute control over the IDT and any corresponding data, including the right to alter the testimonies any way they see fit [Mar+22]. Granting or withholding access, as well as monetization, is also at their sole discretion. For many IDTs, access can be gained via a subscription which is associated with recurring fees. This serves to cover the expenses of providing IDTs as a service (e.g., web hosting) and safeguard against improper use. However, the delineation between proper and improper use is likewise internally decided by the same institution, which can lead to conflicts of interest and unilateral dependencies for customers and users.

As we discovered in Section 3.2.3.2, even IDTs on historical topics can carry an additional personal aspect for their grieving bereaved. This prompts numerous legal and ethical questions and considerations regarding digital afterlife services:

- Should grandchildren be guaranteed a copy of their grandfather's IDT and recordings for personal use? Which degree of kinship entitles one to have access to a loved one's life story?
- What becomes of an IDT and its associated recordings if their provider goes bankrupt?
- Should the wishes of relatives and friends be considered when an IDT of a recently deceased is displayed publicly or for promotional purposes?
- What kinds of commercialization of oral history are acceptable?

While current practices voluntarily employ ample goodwill towards the bereaved, guidelines, regulatory measures, fiduciaries [SP20], and open-source frameworks for creating IDTs are necessary to ensure this remains the case.

### 7.2.2 What is authenticity in the context of IDTs?

Unlike ECAs offering AI-generated output, IDTs derive authenticity in accordance with our requirements from the exclusive use of recordings of genuine testimonies and reactions of real people [For22].

However, despite emphasizing that the collected data are not processed in any way [MCW17], color grading, editing, and trimming are customary and frequently necessary when extracting individual response clips from the continuously recorded raw material [BBG19]. Consequently, the definition of authenticity in the context of IDTs and the tolerance of deviations remains unclear [BG20]. This leads to numerous currently unresolved ethical dilemmas and uncertainties:

- Would an answer that was extracted from a different, longer answer be less authentic?
- Would shorter, more digestible responses or removing factual inaccuracies [Mar+22] undermine authenticity?
- How many seconds of silent pondering at the beginning of an answer can be trimmed before the answer is considered inauthentic?
- Does ML-based re-processing of visual qualities of the recordings [HKD21] impair their authenticity?

Uncertainty and subjective decisions are also inherent to the IDT's AI-based question-answer matching system, whose training relies on intervention by humans who evaluate the correctness as well as the authenticity of a match [SU21; Dud21]. Due to the complexity of building a viable Natural Language Processing system from the ground up, current implementations of IDTs instead incorporate pre-existing, proprietary software or services for Automatic Speech Recognition [Tra+15a; Dud21; SU21] or Natural Language Understanding [Dud21; BD21]. However, the use of closed-source software limits the traceability, transparency, and explainability [CB23] of the processing steps undertaken by the respective system components [BG20; SU21].

Consequently, there is a need for defining authenticity of IDTs (and AI-based digital afterlife applications in general) as well as methods and standards for verifying or evaluating said authenticity.

### 7.2.3 How can other digital testimonies or artifacts be incorporated?

IDTs can benefit from including other digital artifacts to provide users with visual aids or valuable context. These can include digital versions of personal sentimental possessions, geographical maps, photographs, books, or virtual recreations of historical sites [SKF12; Jon+23]. However, the seamless and immersive inclusion of such digital artifacts has proven difficult as they can distract from the conversation or conflict with the display method of the digital witness [Mar+22].

Furthermore, digital artifacts worthy of inclusion can also be other interactive testimonies. Incorporating other interactive testimonies could diversify the presented perspectives by allowing users to ask multiple people the same question and contrast their respective answers.

It is therefore necessary to research and develop concepts on how IDTs can be combined with other digital content. This entails examining how individuals choose and shape their digital legacy and the emotional bonds users form with inherited digital information [KB08]. The identified concepts can append our list of requirements for IDTs, particularly the requirement modularity, as well as our conceptual model.

### 7.2.4 How should data be elicited?

IDTs require extensive data on or by the featured individuals. These sets of data may serve as training data or even directly constitute the actions or reactions displayed.

The creation of high-quality IDTs is an expensive and laborious undertaking that requires several specialists as well as considerable financial resources. A substantial part of the creation process is eliciting and recording the answers of contemporary witnesses. The recording process itself usually aims to gather more than 1000 responses over five consecutive days while retaining cinematic continuity across answers and filming days [Tra+15b; MCW15; MCW17; BBG19]. Due to the accompanying increase in costs and effort, multiple disconnected recording sessions with intermediate empirical analyses of gaps in answer coverage [Art+15] are a scant exception. Consequently, the pool of available answers is determined by the set of answers captured during the first (and often only) recording session. This high-pressure and exhausting environment can entice questionable or outright unethical behavior, such as drilling down on distressing topics to receive "better" content or recklessly pushing for gathering as many answers as possible over the limited time for which the studio is rented. This not only jeopardizes the physical or mental well-being of the people involved (e.g., Holocaust survivors or similarly vulnerable seniors) but also the integrity of the answers provided.

At the time of writing, details on the recording processes and other data collection methods for IDTs are underreported. Expedient referencable ethics standards could thus help mitigate this issue.

### 7.2.5 How to further prepare for future developments?

As previously discussed in Section 2.2.6, IDTs need to adapt to time-related changes in order to continue to be viable in the future. Consequently, employing multi-lifespan design thinking [Yoo+16] is necessary to envision the context of use, technical conditions, and cultural circumstances many years into the future.

Current IDT provisions for time-related changes address two areas: processing user input and visual output quality. The components responsible for processing user input need to be continuously adjusted to keep pace with changes in the languages used as an input medium, e.g., more inclusive terms [SU21]. The visual output quality of IDTs is 'future-proofed' by capturing higher levels and resolutions than presently required. This allows future upgrades to the output and keeping pace with upcoming display technologies without resorting to ML-based generative methods. Cues about the dated nature of the pre-recorded content itself, on the other hand, are deliberately not hidden and left unchanged.

However, there is no guarantee that no other aspects will become susceptible to time-related changes. Similarly, it is unclear how to proceed with upgrading the IDT output once the recorded qualities have been exhausted. Consequently, it is essential to continue to explore and identify other potentially time-sensitive components or properties and prepare for their substitution to enable durable thanatosensitive IDT designs.

### 7.2.6 How should IDTs adapt to users?

While we have contributed substantial empirical research and insights into the advantages and disadvantages of different ways of visually displaying IDTs, there are many more input and output modalities that still need to be considered, investigated, and analyzed. This includes technologies that do not exist yet or are not viable yet. Therefore, research into which combinations of input and output options are most advantageous for which user groups is still ongoing. Consequently, it also remains to be determined which sets of input and output options need to be supported to provide all necessary combinations.

This is linked to future research into how interactions with IDTs can best be offered and adapted to users. One possible approach could be dynamic customization based on an analysis of cues and characteristics of users and their inputs. Heuristics and neural networks could be used to automatically adapt the IDT output, for example by linking the subtitles to the language of the input or recognizing and taking into account assistive technologies such as vision or hearing aids.

### 7.2.7 How should IDTs be integrated into educational concepts?

Within our research, we found a broad spectrum and considerable differences in the way IDTs are integrated into the educational concepts of their respective exhibition sites. These integrations of IDTs range from unconnected stand-alone offers to ingrained embedding in museum visits or tours. We encountered similar variances in the background and training of on-site educators, such as teachers or museum staff.

The integration of IDTs into the curricula of schools, museums, or other places of learning could benefit from the further development of standards, theories, and guidelines [WM23], e.g., lessons on working with oral history and historical sources in general. This is also accompanied by the need for underpinning empirical data on the impact of each concept on learning success as well as the critical evaluation and reflection of current integrations into educational concepts.

### 7.2.8 How does the learning with IDTs compare to conversations with human contemporary witnesses?

The scientific rationale for the capability of IDTs to provide an alternative to real conversations with human witnesses is robust, but also largely hypothetical. At the time of writing, there is a categorical lack of empirical comparisons between conversations with IDTs and conversations with human contemporary witnesses. Such studies are necessary to understand how and, if applicable, why user behaviors might differ:

- Do they ask about the same topics?
- Do they phrase their questions differently?
- Is interacting with an IDT instead of a human contemporary witness either inhibiting or reassuring?
- Are there differences in the learning qualities or methods?

These types of future research deserve heightened attention, especially as long as such direct comparisons are still possible.

# Chapter 8

# Conclusions

*"The purpose of memory is not simply to preserve the past, it is to protect the future."*
— Barack Obama*, May 7th, 2014* [29]

In this thesis, we investigated **RQ:** "How can Interactive Digital Testimonies be designed to preserve conversations with contemporary witnesses?". We divided our investigation into four subquestions, which we revisit in the following:

**SQ-1:** What are the requirements to digitally preserve interactive conversations with contemporary witnesses?

- Due to thematic proximity and overlaps in scope and content, we investigated **SQ-1** in tandem with **SQ-2**.
- In Section 2.2 we systematically compiled, reviewed, and analyzed the literature on IDTs to examine the requirements and properties of IDTs. This constitutes contribution **C-1**.
- We identified that R1: Authenticity, R2: Conversational Input, R3: Lifelike Output, R4: Accessibility, R5: Revisability, and R6: Self-sufficiency are requirements for IDTs. A detailed description can be found in Section 2.5.
- We conducted a comprehensive semi-structured qualitative study of IDT experts and stakeholders, which we detail in Chapter 3. This constitutes contribution **C-2**. Section A of the Appendix provides additional details on this study.
- Based on our findings, we revised and appended our list of requirements for IDTs to include the requirements R7: Adaptability and R8: Self-contained. The revised list of requirements is presented in Section 3.4.

**SQ-2:** What are the associated practices or shortcomings of current IDT implementations?

- In Section 2.4 we examined the extent to which previous IDT design models do not meet the identified requirements.
- As part of our expert interviews (contribution **C-2**), we identified three overarching themes: The difficulties encountered by users when engaging with IDTs, the possibilities and limitations in accessing IDTs, and the unexpected emotional roles that IDTs can assume. We report our findings in Section 3.2.

**SQ-3:** How can IDTs be designed to fulfill the identified requirements?

- Building on our requirement analysis, we created a conceptual model of IDT designs in Chapter 4. This constitutes contribution **C-3**.

[29]President Obama Speaks at the USC Shoah Foundation Dinner: https://www.youtube.com/watch?v=WwBIzWN6u00#t=529

- We found that a modular, instantiable IDT design consisting of the components Input Device, Preprocessor, Matching System, Database of Recordings, Archive, and Display can implement and fulfill the previously identified list of requirements.
- In Chapter 5, we outlined how the LediZ IDTs serve as a proof of concept of our conceptual model of IDT designs. This constitutes contribution **C-4**.
- In the LediZ project, we use a three-pronged implementation of IDTs: A stationary high-fidelity instance with a focus on immersiveness, a transportable instance that aims to retain as much as possible of the stationary instance, and an online instance. A centralized system provides a preprocessor, a matching system, and a database of recordings, which also serve as a ground truth for the instantiations. An archive stores the unaltered recorded data, the processed versions of which populate the database of recordings within the centralized system. To ensure adaptability and modularity, we employ technical standards, standardized interfaces, and standardized data formats.
- Our approach implements all the components and overarching properties of our conceptual model and consequently addresses each of the previously identified requirements.

**SQ-4:** How does the type of visual output modality affect how users experience the interaction with the implemented IDT?

- In Chapter 6, we presented an empirical evaluation of the implemented IDTs in general and their visual modality in particular through two distinct mixed-methods user studies. This constitutes contribution **C-5**. Section B and Section C of the Appendix provide additional details on both studies.
- We found that audio-visual 3D IDTs are perceived as more authentic and engaging, however, discomfort caused by 3D glasses undermines advantages over experiences with audio-visual 2D IDTs. 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 reaffirm the need for IDTs to be able to dynamically adapt their interaction conditions to the user.

In Chapter 7, we outline future directions for research and work that can build on and benefit from our findings, insights, and contributions.

We found IDTs to be an expedient approach for preserving conversations with contemporary witnesses. However, these systems necessitate — and merit — significantly more scientific attention from a broad range of research disciplines, alongside the further development of prototypes, iterations, and refinements, to one day be able to comprehensively capture and reproduce the full depth of extensive interactions with real human beings. We look forward to the future discussions, findings, and technical advances that will enable IDTs to move closer to this goal, in the hope that these will help to democratize and perpetuate access to knowledge in general and to intangible cultural heritage in particular.

# Acknowledgements

Like any and all challenging undertakings, this doctoral project and its associated successes had many architects. I owe a debt of deep gratitude to several individuals*.

Chiefly and supremely to Lara Hirschbeck, who unwaveringly supported me throughout this journey, who stood and continues to stand by my side, rain or shine, who helped me grow into who I am today, and without whom truly none of this would have been possible. Similarly, to Tanja Reichert, Felix Ecker, Benjamin Fritzsche, Luisa Fritzsche, Elisabeth Hupel, Evelyn Reiter, Valentin Baumann, Freddy Baumann, the late Gerhard Winkler, and Edeltraud (Mona) Kolb, who always had confidence in me and for whom anything less than a doctorate would have been a surprise.

To Thomas Odaker, Simone Müller, Elisabeth Mayer, Lea Weil, Maxime Pedrotti, Daniel Huber, Rainer Oesmann, Jutta Dreer, and Kris Weinand of the extraordinary VVR team at the LRZ as well as Sabine Osorio and Severine Müller (also at the LRZ but in other, equally extraordinary teams), with whom I may have started out as mere colleagues, but who have become so much more to me. For their myriad contributions, be it conducive discussions or emotional or manual-technical support. For fostering such a blissful, considerate, and courteous atmosphere and setting exceptionally high standards for legitimate teamwork and true leadership; standards that future coworkers or supervisors may struggle with.

To my colleagues and friends at the LediZ project with whom I had the joy of exploring and researching the field of IDTs together and with whom I had to overcome many a challenge in faraway and unfamiliar places. This illustrious group includes the Holocaust survivors Eva Umlauf and Abba Naor, and fellow researchers Anja Ballis, Markus Gloe, Fabian Heindl, Ernst Hüttl, Frauke Teichmann, Florian Duda, and Alessa Diehl.

To my invaluable DnD group of lovable lunatics who provided a much-needed and longed-for distraction each and every Thursday, and particularly to Hazel Gibson for her auspicious suggestions, recommendations, and consolation.

To all my constructive and rewarding co-authors, including Jens Weismüller, Wolfgang Kurtz, and specifically the late Alexander von Ramm, with whom I had the all too brief pleasure of sharing an office before his harrowing, untimely passing and whose shuffling down the hallway I miss dearly.

To Saad Elbeleidy, Aisling Kelliher, Sophia Grundner-Culemann, Johannes Albert-von der Gönna, Tobias Weber, Doris Hausen, Christoph Anthes, Sylvia Rothe, Jule Ziegler, Kerstin Brumund, Florian Liedtke, and the late Heinrich Hußmann for their encouragement, support, insights, and advice for and throughout the journey of a doctoral student.

To the late Kia Hays and Kim Simon, who welcomed us so kindly and cordially at the USC Shoah Foundation, who made our stay stimulating and informative, and who died far before their time.

To the brilliant students I have had the privilege of mentoring, supervising, or working with over the years and from whom I have sometimes learned more than they might have from me.

To all the interviewees and study participants, who took a chance on me and graciously granted me some of their valuable time.

To Alexandra Elbakyan and the late Aaron Swartz for their sacrifices in the name of open science.

To all four-legged friends — if they somehow manage to read this —, whether familiar or unfamiliar, past or present, and to those who are committed to animal welfare in general.

To Don Rosa, whose magnum opus “The Life and Times of Scrooge McDuck” inspired me and instilled in me the importance of tenacity and principles in the face of adversity.

Lastly, to the distinguished group of composers and musicians whose soothing melodies strengthened and sharpened my focus: Erik Satie, Claude Debussy, Sergei Rachmaninoff, Ludwig van Beethoven, Giuseppe Verdi, Tuomas Holopainen, Queen, Pink Floyd, and Paweł Zadrożniak, to name but a few. A life without their music would be possible, but pointless.

*Individuals who belong to more than one of the groups listed are eligible for and entitled to repeated gratitude.

# Bibliography


[AG19] Gabi Arrigoni and Areti Galani. "Digitally enhanced polyvocality and reflective spaces : Challenges in sustaining dialogue in museums through digital technologies". In: *European Heritage, Dialogue and Digital Practices*. Routledge, 2019, pp. 37–61. DOI: 10.4324/9780429053511-3.

[AG21] Gabi Arrigoni and Areti Galani. "Recasting witnessing in museums: digital interactive displays for dialogic remembering". In: *International Journal of Heritage Studies* 27.2 (2021), pp. 250–264. DOI: 10.1080/13527258.2020.1795909.

[Ala18] Wafaa Abdullah Alamri. "Effectiveness of Qualitative Research Methods: Interviews and Diaries". In: *International Journal of English and Cultural Studies* 2.1 (2018), p. 65. ISSN: 2575-8101. DOI: 10.11114/ijecs.v2i1.4302.

[Alc88] Linda Alcoff. "Cultural Feminism versus Post-Structuralism: The Identity Crisis in Feminist Theory". In: *Signs: Journal of Women in Culture and Society* 13.3 (1988), pp. 405–436. DOI: 10.1086/494426.

[Ale21] Neta Alexander. "Obsolescence, Forgotten: "Survivor Holograms", Virtual Reality, and the Future of Holocaust Commemoration". In: *Cinergie – Il Cinema e le altre Arti* 10.19 (2021), pp. 57–68. ISSN: 2280-9481. DOI: 10.6092/issn.2280-9481/12205.

[Alj+19] Hussain M. Aljaroodi, Marc T. P. Adam, Raymond Chiong, and Timm Teubner. "Avatars and Embodied Agents in Experimental Information Systems Research: A Systematic Review and Conceptual Framework". In: *Australasian Journal of Information Systems* 23 (2019). ISSN: 1449-8618. DOI: 10.3127/ajis.v23i0.1841. URL: http://journal.acs.org.au/index.php/ajis/article/view/1841.

[AMD22] Luis Almeida, Paulo Menezes, and Jorge Dias. "Telepresence Social Robotics towards Co-Presence: A Review". In: *Applied Sciences* 12.11 (2022), p. 5557. ISSN: 2076-3417. DOI: 10.3390/app12115557. URL: https://www.mdpi.com/1656246.

[Amo23] A.D. Amorosi. "KISS Says Farewell at MSG, Then Passes the Torch to Avatar Successors". In: *Variety* (2023). URL: https://variety.com/2023/music/concert-reviews/kiss-final-concert-review-madison-square-garden-avatars-1235819744/ (visited on 03/27/2024).

[And11] Elisabeth André. "Design and Evaluation of Embodied Conversational Agents for Educational and Advisory Software". In: *Gaming and Simulations: Concepts, Methodologies, Tools and Applications*. Hershey, PA, USA: IGI Global, 2011, pp. 668–686. DOI: 10.4018/978-1-60960-195-9.ch306. URL: https://www.igi-global.com/chapter/gaming-simulations-concepts-methodologies-tools/49411.

[Art+14] Ron Artstein, David Traum, Oleg Alexander, Anton Leuski, Andrew Jones, Kallirroi Georgila, Paul Debevec, William Swartout, Heather Maio, and Stephen Smith. "Time-offset interaction with a holocaust survivor". In: *Compilation publication of IUI'14 proceedings & IUI'14 companion*. Ed. by Tsvi Kuflik. New York, NY: ACM, 2014, pp. 163–168. ISBN: 9781450321846. DOI: 10.1145/2557500.2557540.

[Art+15] Ron Artstein, Anton Leuski, Heather Maio, Tomer Mor-Barak, Carla Gordon, and David Traum. "How many utterances are needed to support time-offset interaction?" In: *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*. 2015. URL: https://www.aaai.org/ocs/index.php/flairs/flairs15/paper/viewpaper/10442.

[Bac20] Francis Bacon. *Novum Organum, sive Indicia Vera de Interpretatione Naturae [The New Organon, or True Directions Concerning the Interpation of Nature]*. Ed. by Lisa Jardine and Michael Silverthorne. Translation published 2000. Cambridge University Press, 1620. ISBN: 9781139427289.

[Bai+06] Jeremy N. Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. "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". In: *Presence: Teleoperators and Virtual Environments* 15.4 (2006), pp. 359–372. DOI: 10.1162/pres.15.4.359.

[Bal19] Anja Ballis. ""With a little antisemitism, you could live!" A case study on framing a Holocaust survivor's living testimony". In: *Theo-Web* 18.1 (2019). DOI: 10.23770/tw0085.

[Bal21a] Anja Ballis. "Interaktive 3D-Zeugnisse von Holocaust-Überlebenden im Deutschunterricht – Theoretische Rahmung, empirische Exploration und disziplinäre Zielhorizonte". In: *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie*. Ed. by Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann. Vol. 1. Eckert. Dossiers. Braunschweig, Germany: Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung, 2021, pp. 83–110.

[Bal21b] Anja Ballis. "Memories and Media: Pinchas Gutter's Holocaust Testimonies". In: *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie*. Ed. by Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann. Vol. 1. Eckert. Dossiers. Braunschweig, Germany: Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung, 2021, pp. 147–166.

[Bal22] Anja Ballis. "The Impact of Digitization on Tour Guiding". In: *Tour Guides at Memorial Sites and Holocaust Museums*. Ed. by Anja Ballis. Holocaust Education – Historisches Lernen – Menschenrechtsbildung. Wiesbaden: Springer Fachmedien Wiesbaden, 2022, pp. 195–212. ISBN: 978-3-658-35817-4. DOI: 10.1007/978-3-658-35818-1_11.

[BB21] Ben Beaumont-Thomas and Mark Brown. "Abba reunite for Voyage, first new album in 40 years". In: *The Guardian* (2021). URL: https://www.theguardian.com/music/2021/sep/02/abba-reunite-for-voyage-first-new-album-in-40-years (visited on 03/27/2024).

[BBG19] Anja Ballis, Michele Barricelli, and Markus Gloe. "Interaktive digitale 3-D-Zeugnisse und Holocaust Education – Entwicklung, Präsentation und Erforschung". In: *Holocaust Education Revisited*. Ed. by Anja Ballis and Markus Gloe. Holocaust Education – Historisches Lernen – Menschenrechtsbildung. Wiesbaden: Springer Fachmedien Wiesbaden, 2019, pp. 403–436. ISBN: 9783658242053. DOI: 10.1007/978-3-658-24205-3_22.

[BC06] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative Research in Psychology* 3.2 (2006), pp. 77–101. DOI: 10.1191/1478088706qp063oa.

[BC22] Virginia Braun and Victoria Clarke. "Conceptual and design thinking for thematic analysis". In: *Qualitative Psychology* 9.1 (2022), pp. 3–26. ISSN: 2326-3598. DOI: 10.1037/qup0000196.

[BD21] Anja Ballis and Florian Duda. "Zwischen Mensch und Maschine?! – Schüler*innen befragen das interaktive 3D-Zeugnis eines Holocaust-Überlebenden". In: *Mitteilungen des Deutschen Germanistenverbandes* 68.3 (2021), pp. 284–291. DOI: 10.14220/mdge.2021.68.3.284.

[BDG21] Anja Ballis, Florian Duda, and Markus Gloe. "Lernen mit interaktiven 3D-Zeugnissen – Perspektiven aus dem und auf das Münchner Projekt LediZ". In: *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie*. Ed. by Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann. Vol. 1. Eckert. Dossiers. Braunschweig, Germany: Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung, 2021, pp. 2–24.

[Ber22] Nora Berner. “Life Stories as Memory Carriers”. In: *Remembrance – Responsibility – Reconciliation*. Ed. by Lothar Wigger and Marie Dirnberger. Kindheit – Bildung – Erziehung. Philosophische Perspektiven. Berlin, Heidelberg: Springer Berlin Heidelberg and Imprint J.B. Metzler, 2022, pp. 141–155. ISBN: 978-3-662-64184-2. DOI: 10.1007/978-3-662-64185-9_10.

[BF07] Aaron Beim and Gary Alan Fine. “Trust in Testimony: The Institutional Embeddedness of Holocaust Survivor Narratives”. In: *European Journal of Sociology* 48.1 (2007), pp. 55–75. ISSN: 0003-9756. DOI: 10.1017/s000397560700029x.

[BG20] Anja Ballis and Markus Gloe. “Interactive 3D Testimonies of Holocaust Survivors in German language”. In: *Holocaust Education Revisited*. Ed. by Markus Gloe and Anja Ballis. Holocaust Education – Historisches Lernen – Menschenrechtsbildung. Wiesbaden: Springer Fachmedien Wiesbaden, 2020, pp. 343–368. ISBN: 978-3-658-24207-7. DOI: 10.1007/978-3-658-24207-7_21.

[BHB03] Frank Biocca, Chad Harms, and Judee K. Burgoon. “Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria”. In: *Presence: Teleoperators and Virtual Environments* 12.5 (2003), pp. 456–480. DOI: 10.1162/105474603322761270.

[BKV22] Andreas Butz, Antonio Krüger, and Sarah Theres Völkel. *Mensch-Maschine-Interaktion*. 3. Auflage. De Gruyter Studium. Berlin and Boston: De Gruyter Oldenbourg, 2022. ISBN: 9783110753325. DOI: 10.1515/9783110753325.

[Bla+17] María José Blanca Mena, Rafael Alarcón Postigo, Jaume Arnau Gras, Roser Bono Cabré, and Rebecca Bendayan. “Non-normal data: Is ANOVA still a valid option?” In: *Psicothema* 29.4 (2017), pp. 552–557. DOI: 10.7334/psicothema2016.383. URL: https://diposit.ub.edu/dspace/handle/2445/122126.

[BLM09] Alexander Bogner, Beate Littig, and Wolfgang Menz. “Introduction: Expert Interviews — An Introduction to a New Methodological Debate”. In: *Interviewing experts*. Ed. by Alexander Bogner, Beate Littig, and Wolfgang Menz. Research methods series. Basingstoke England and New York: Palgrave Macmillan, 2009, pp. 1–13. ISBN: 9780230244276. DOI: 10.1057/9780230244276_1. URL: https://link.springer.com/chapter/10.1057/9780230244276_1.

[Bru+24] Jed R. Brubaker, Meredith Ringel Morris, Dylan Thomas Doyle, Casey Fiesler, Martin Gibbs, and Joanna McGrenere. “AI and the Afterlife”. In: *CHI’24*. Ed. by Florian Mueller and Penny Kyburz. ACM Digital Library. New York, New York: The Association for Computing Machinery, 2024, pp. 1–5. ISBN: 9798400703317. DOI: 10.1145/3613905.3636321.

[Brü18] Christina Isabel Brüning. “Hologramme von Überlebenden in einer sich diversifizierenden Gesellschaft?” In: *Totalitarianism and Democracy* 15.2 (2018), pp. 219–232. ISSN: 1612-9008. DOI: 10.13109/tode.2018.15.2.219.

[Brü19] Christina Isabel Brüning. “Holocaust Education in Multicultural Classrooms. Some Insights into an Empirical Study on the Use of Digital Survivor Testimonies”. In: *Holocaust Education Revisited*. Ed. by Anja Ballis and Markus Gloe. Holocaust Education – Historisches Lernen – Menschenrechtsbildung. Wiesbaden: Springer Fachmedien Wiesbaden, 2019, pp. 391–402. ISBN: 9783658242053. DOI: 10.1007/978-3-658-24205-3_21.

[BS21] Anja Ballis and Lisa Schwendemann. “‘In any case, you believe him one hundred percent, everything he says.’: Trustworthiness in Holocaust survivor talks with high school students in Germany”. In: *Holocaust Studies* (2021), pp. 1–30. DOI: 10.1080/17504902.2021.1915016.

[BSv20] Hanneke Bartelds, Geerte M. Savenije, and Carla van Boxtel. “Students’ and teachers’ beliefs about historical empathy in secondary history education”. In: *Theory & Research in Social Education* 48.4 (2020), pp. 529–551. ISSN: 0093-3104. DOI: 10.1080/00933104.2020.1808131.

[BW14] Adam Brown and Deb Waterhouse-Watson. “The Future of the Past: Digital Media in Holocaust Museums”. In: *Holocaust Studies* 20.3 (2014), pp. 1–32. DOI: 10.1080/17504902.2014.11435374.

[BWT17] Christiane Bertram, Wolfgang Wagner, and Ulrich Trautwein. “Learning Historical Thinking With Oral History Interviews: A Cluster Randomized Controlled Intervention Study of Oral History Interviews in History Lessons”. In: *American Educational Research Journal* 54.3 (2017), pp. 444–484. DOI: 10.3102/0002831217694833.

[Byr22] David Byrne. “A worked example of Braun and Clarke’s approach to reflexive thematic analysis”. In: *Quality & Quantity* 56.3 (2022), pp. 1391–1412. ISSN: 1573-7845. DOI: 10.1007/s11135-021-01182-y. URL: https://link.springer.com/article/10.1007/s11135-021-01182-y.

[Cal+08] Ben Caldwell, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. *Web content accessibility guidelines (WCAG) 2.0*. 2008. URL: https://www.w3.org/TR/WCAG20/.

[Cas00] Justine Cassell. “Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents”. In: *Embodied conversational agents*. Ed. by Justine Cassell, Tim Brickmore, Lee Campbell, Hannes Vilhjálmsson, and Hao Yan. Cambridge, Mass.: MIT Press, 2000, pp. 1–27. ISBN: 9780262032780. URL: https://web.media.mit.edu/~cynthiab/readings/cassell-eca-00.pdf.

[CB23] Tara Capel and Margot Brereton. “What is Human-Centered about Human-Centered AI? A Map of the Research Landscape”. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2023. DOI: 10.1145/3544548.3580959.

[Cha+22] Ishita Chakraborty, Khai Chiong, Howard Dover, and K. Sudhir. *AI and AI-Human Based Salesforce Hiring Using Interview Videos*. 2022. DOI: 10.2139/ssrn.4137872.

[Cha90] Lucy Chabot. “Nixon Library Technology Lets Visitors ’Interview’ Him”. In: *Los Angeles Times (July 21st)* (1990). URL: https://www.latimes.com/archives/la-xpm-1990-07-21-mn-346-story.html (visited on 09/21/2023).

[CL19] Emna Chérif and Jean-François Lemoine. “Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant’s voice”. In: *Recherche et Applications en Marketing (English Edition)* 34.1 (2019), pp. 28–47. DOI: 10.1177/2051570719829432.

[CLL23] Yunjae J. Choi, Minha Lee, and Sangsu Lee. “Toward a Multilingual Conversational Agent: Challenges and Expectations of Code-mixing Multilingual Users”. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2023. DOI: 10.1145/3544548.3581445.

[Com+89] D. E. Comer, David Gries, Michael C. Mulder, Allen Tucker, A. Joe Turner, and Paul R. Young. “Computing as a discipline”. In: *Communications of the ACM* 32.1 (1989), pp. 9–23. ISSN: 0001-0782. DOI: 10.1145/63238.63239.

[Con20] Randy Connolly. “Why computing belongs within the social sciences”. In: *Communications of the ACM* 63.8 (2020), pp. 54–59. ISSN: 0001-0782. DOI: 10.1145/3383444.

[Cze+97] Marek Czernuszenko, Dave Pape, Daniel Sandin, Tom DeFanti, Gregory L. Dawe, and Maxine D. Brown. “The ImmersaDesk and Infinity Wall projection-based virtual reality displays”. In: *ACM SIGGRAPH Computer Graphics* 31.2 (1997), pp. 46–49. DOI: 10.1145/271283.271303.

[Dah+16] Salam Daher, Kangsoo Kim, Myungho Lee, Andrew Raij, Ryan Schubert, Jeremy Bailenson, and Greg Welch. “Exploring social presence transfer in real-virtual human interaction”. In: *2016 IEEE Virtual Reality (VR)*. Ed. by Tobias Höllerer, Victoria Interrante, Anatole Lécuyer, and Evan Suma. Piscataway, NJ, USA: IEEE, 2016, pp. 165–166. DOI: 10.1109/vr.2016.7504705.

[Die+22] Stephan Diederich, Alfred Benedikt Brendel, Stefan Morana, and Lutz Kolbe. “On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research”. In: *Journal of the Association for Information Systems* 23.1 (2022), pp. 96–138. ISSN: 1536-9323. DOI: 10.17705/1jais.00724.

[Dix+04] Alan Dix, Janet Finlay, Gregory D. Abowd, and Russell Beale. *Human-computer interaction*. Third. Harlow: Pearson Prentice-Hall, 2004. ISBN: 9780130461094.

[DL86] Richard L. Daft and Robert H. Lengel. "Organizational Information Requirements, Media Richness and Structural Design". In: *Management Science* 32.5 (1986), pp. 554–571. ISSN: 0025-1909. DOI: 10.1287/mnsc.32.5.554.

[Dud21] Florian Duda. ""Was war oder ist Ihre schönste, tollste und angenehmste Kindheitserinnerung?" – Ein sprachwissenschaftlicher Ansatz zur Machine-Learning-Datengenerierung". In: *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie*. Ed. by Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann. Vol. 1. Eckert. Dossiers. Braunschweig, Germany: Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung, 2021, pp. 43–62. DOI: 10.5282/ubm/epub.75160.

[EB13] Jason Endacott and Sarah Brooks. "An Updated Theoretical and Practical Model for Promoting Historical Empathy". In: *Social Studies Research and Practice* 8.1 (2013), pp. 41–58. DOI: 10.1108/SSRP-01-2013-B0003.

[ED13] Aaron C. Elkins and Douglas C. Derrick. "The Sound of Trust: Voice as a Measurement of Trust During Interactions with Embodied Conversational Agents". In: *Group Decision and Negotiation* 22.5 (2013), pp. 897–913. ISSN: 1572-9907. DOI: 10.1007/s10726-012-9339-x. URL: https://link.springer.com/article/10.1007/s10726-012-9339-x.

[Er-+17] Adil Er-Rady, R. Faizi, R. Oulad Haj Thami, and H. Housni. "Automatic sign language recognition: A survey". In: *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, 2017. DOI: 10.1109/atsip.2017.8075561.

[Fei+19] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. "A Taxonomy of Social Cues for Conversational Agents". In: *International Journal of Human-Computer Studies* 132 (2019), pp. 138–161. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2019.07.009.

[Fei59] Louis Fein. "The role of the University in computers, data processing, and related fields". In: *Communications of the ACM* 2.9 (1959), pp. 7–14. ISSN: 0001-0782. DOI: 10.1145/368424.368427.

[Fli18] Uwe Flick. *An introduction to qualitative research*. Sixth edition. Los Angeles: SAGE, 2018. ISBN: 9781526464217.

[For22] Rebecca Ford. "Teaching Trauma Without Traumatizing". In: *The iJournal: Student Journal of the Faculty of Information* 7.3 (2022), pp. 60–69-60–69. ISSN: 2561-7397. DOI: 10.33137/ijournal.v7i3.39326. URL: https://theijournal.ca/index.php/ijournal/article/view/39326.

[Fou23] Sofia Platform Foundation. *Histories from Belene*. 2023. URL: https://belene.camp/en/ (visited on 02/27/2024).

[Fro18] Paul Frosh. "The mouse, the screen and the Holocaust witness: Interface aesthetics and moral response". In: *New Media & Society* 20.1 (2018), pp. 351–368. DOI: 10.1177/1461444816663480.

[Gal14] Phil Gallo. "Michael Jackson Hologram Rocks Billboard Music Awards: Watch & Go Behind the Scenes". In: *Billboard* (2014). URL: https://www.billboard.com/music/music-news/michael-jackson-hologram-billboard-music-awards-6092040/ (visited on 03/27/2024).

[Gam21] Cayo Gamber. "Emerging Technologies and the Advent of the Holocaust "Hologram"". In: *Emerging Technologies and the Digital Transformation of Museums and Heritage Sites*. Ed. by Maria Shehade and Theopisti Stylianou-Lambert. Communications in Computer and Information Science. Springer, Cham, 2021, pp. 217–231. ISBN: 978-3-030-83646-7. DOI: 10.1007/978-3-030-83647-4_15.

[Gav91] William W. Gaver. "Technology affordances". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Ed. by Scott P. Robertson. New York, NY: ACM, 1991, pp. 79–84. ISBN: 0897913833. DOI: 10.1145/108844.108856.

[GLL99] Joakim Gustafson, Nikolaj Lindberg, and Magnus Lundeberg. "The August spoken dialogue system". In: *6th European Conference on Speech Communication and Technology*. 1999, pp. 1151–1154. DOI: 10.21437/Eurospeech.1999-268. URL: https://www.isca-speech.org/archive/eurospeech_1999/gustafson99_eurospeech.html.

[Glo21] Markus Gloe. "Digital Interactive 2D/3D Testimonies in Holocaust Museums in the United States and Europe". In: *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie*. Ed. by Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann. Vol. 1. Eckert. Dossiers. Braunschweig, Germany: Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung, 2021, pp. 130–146.

[Gof63] Erving Goffman. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York: Simon & Schuster, 1963. ISBN: 9781439108697.

[Gol+22] Ariel Goldman, Cindy Espinosa, Shivani Patel, Francesca Cavuoti, Jade Chen, Alexandra Cheng, Sabrina Meng, Aditi Patil, Lydia B. Chilton, and Sarah Morrison-Smith. "QuAD: Deep-Learning Assisted Qualitative Data Analysis with Affinity Diagrams". In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. New York, NY, USA: ACM, 2022. DOI: 10.1145/3491101.3519863.

[Got+18] Daniel Gotsch, Xujing Zhang, Timothy Merritt, and Roel Vertegaal. "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*. New York, NY, USA: ACM, 2018. DOI: 10.1145/3173574.3174096.

[Gra+02] J. Gratch, J. Rickel, E. Andre, J. Cassell, E. Petajan, and N. Badler. "Creating interactive virtual humans: some assembly required". In: *IEEE Intelligent Systems* 17.4 (2002), pp. 54–63. ISSN: 1541-1672. DOI: 10.1109/mis.2002.1024753.

[Gro16] Kory Grow. "Ronnie James Dio Hologram Debuts at German Metal Festival". In: *Rolling Stone* (2016). URL: https://www.rollingstone.com/music/music-news/ronnie-james-dio-hologram-debuts-at-german-metal-festival-96261/ (visited on 03/27/2024).

[GW20] Antje Gansewig and Maria Walsh. *Zeitzeugen in der schulischen Bildungsarbeit und deren Potenzial zur Wertevermittlung – Einsatz und empirische Befundlage*. Vol. 74. Jahrgang / 2020. Peter Lang Academic Publishing Group, 2020. DOI: 10.3726/PR052020.0052.

[HB98] Karen Holtzblatt and Hugh Beyer. *Contextual design: Defining customer-centered systems*. Interactive Technologies. San Francisco, Calif: Morgan Kaufmann, 1998. ISBN: 9780080503042.

[Hec11] Douglas D. Heckathorn. "Comment: Snowball versus Respondent-Driven Sampling". In: *Sociological Methodology* 41.1 (2011), pp. 355–366. ISSN: 0081-1750. DOI: 10.1111/j.1467-9531.2011.01244.x.

[Hei21] Fabian Heindl. "The Role of Narrative Structures and Contextual Information in Digital Interactive 3D Testimonies". In: *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie*. Ed. by Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann. Vol. 1. Eckert. Dossiers. Braunschweig, Germany: Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung, 2021, pp. 111–129. DOI: 10.5282/ubm/epub.75107.

[HH15] Gunnar Harboe and Elaine M. Huang. "Real-World Affinity Diagramming Practices". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2015. DOI: 10.1145/2702123.2702561.

[HIA23] Joandi Hartendorp, Nicole Immler, and Hans Alma. "Multi-perspectivity and the risk of perpetration minimisation in Dutch Holocaust and slavery education". In: *Journal of Curriculum Studies* 55.6 (2023), pp. 700–719. ISSN: 0022-0272. DOI: 10.1080/00220272.2023.2261998.

[HKD21] Loc Huynh, Bipin Kishore, and Paul Debevec. *A New Dimension in Testimony: Relighting Video with Reflectance Field Exemplars*. Apr. 6, 2021. URL: http://arxiv.org/pdf/2104.02773v1.

[HN24] Tomasz Hollanek and Katarzyna Nowaczyk-Basińska. "Griefbots, Deadbots, Postmortem Avatars: on Responsible Applications of Generative AI in the Digital Afterlife Industry". In: *Philosophy & Technology* 37.2 (2024), pp. 1–22. ISSN: 2210-5441. DOI: 10.1007/s13347-024-00744-w.

[Hol91] Arno Holz. *Die Kunst: Ihr Wesen und ihre Gesetze*. W. Issleib (G. Schuhr), 1891.

[Hui+17] Tim Huijgen, Carla van Boxtel, Wim van de Grift, and Paul Holthuis. “Toward Historical Perspective Taking: Students’ Reasoning When Contextualizing the Actions of People in the Past”. In: *Theory & Research in Social Education* 45.1 (2017), pp. 110–144. ISSN: 0093-3104. DOI: 10.1080/00933104.2016.1208597.

[Hui18] Tom Huizenga. “Raising The Dead — And A Few Questions — With Maria Callas’ Hologram”. In: *NPR* (2018). URL: https://www.npr.org/sections/deceptivecadence/2018/11/06/664653353/raising-the-dead-and-a-few-questions-with-maria-callas-hologram (visited on 03/27/2024).

[IN00] Katherine Isbister and Clifford Nass. “Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics”. In: *International Journal of Human-Computer Studies* 53.2 (2000), pp. 251–267. ISSN: 1071-5819. DOI: 10.1006/ijhc.2000.0368.

[Int19] International Organization for Standardization. *ISO 9241-210:2019: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.* Geneva, Switzerland, 2019. URL: https://www.iso.org/standard/77520.html (visited on 09/21/2023).

[Int20] International Organization for Standardization. *ISO/IEC 14496-14:2020: Information technology: Coding of audio-visual objects: Part 14: MP4 file format.* Geneva, Switzerland, 2020. URL: https://www.iso.org/standard/79110.html (visited on 03/18/2024).

[IS03] S. Iwarsson and A. Ståhl. “Accessibility, usability and universal design–positioning and definition of concepts describing person-environment relationships”. In: *Disability and Rehabilitation* 25.2 (2003), pp. 57–66. DOI: 10.1080/dre.25.2.57.66.

[JA18] Jasmine Jones and Mark S. Ackerman. “Co-constructing family memory: Understanding the intergenerational practices of passing on family stories”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA, 2018, pp. 1–13.

[JAA21] JAAQ. *Just Ask A Question. The new mental health platform.* 2021. URL: https://jaaq.org/ (visited on 02/27/2024).

[Jen+08] Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. “Measuring and defining the experience of immersion in games”. In: *International Journal of Human-Computer Studies* 66.9 (2008), pp. 641–661. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2008.04.004.

[JL14] Tero Jokela and Andrés Lucero. “MixedNotes”. In: *Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content & Services.* New York, NY, USA: ACM, 2014. DOI: 10.1145/2676467.2676478.

[Jon+23] Val Jones, Weston Gaylord, Ari Palitz, Inge Auerbacher, Andrea Csabai, Dan Franke, Moira Hamilton, Kia Hays, Andrew Lush, Laiylaly Mandujano, Ruxandra Popescu, Francis Rockwell, Zoe Roellin, Ronen Shaham, and Stephen Smith. “Tell Me, Inge. . . ”. In: *ACM SIGGRAPH 2023 Real-Time Live!* New York, NY, USA: ACM, 2023. DOI: 10.1145/3588430.3597256.

[Kam+19] Dorota Kamińska, Tomasz Sapiński, Sławomir Wiak, Toomas Tikk, Rain Haamer, Egils Avots, Ahmed Helmi, Cagri Ozcinar, and Gholamreza Anbarjafari. “Virtual Reality and Its Applications in Education: Survey”. In: *Information* 10.10 (2019), p. 318. DOI: 10.3390/info10100318. URL: https://www.mdpi.com/554326.

[Kam18] Henrik Kampling. “The Role of Immersive Virtual Reality in Individual Learning”. In: *Proceedings of the 51st Hawaii International Conference on System Sciences.* 2018, pp. 1397–1406. ISBN: 978-0-9981331-1-9. DOI: 10.24251/HICSS.2018.173. URL: https://scholarspace.manoa.hawaii.edu/handle/10125/50060.

[Kan14] Wulf Kansteiner. “Genocide memory, digital cultures, and the aesthetization of violence”. In: *Memory Studies* 7.4 (2014), pp. 403–408. ISSN: 1750-6980. DOI: 10.1177/1750698014542389.

[Käs33] Erich Kästner. *Das fliegende Klassenzimmer.* 1933.

[KB08] David Kirk and Richard Banks. “On the design of technology heirlooms”. In: *SIMTech’08.* 2008.

[KB18] Veton Kepuska and Gamal Bohouta. "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)". In: *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018. DOI: 10.1109/ccwc.2018.8301638.

[Ken+93] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness". In: *The International Journal of Aviation Psychology* 3.3 (1993), pp. 203–220. DOI: 10.1207/s15327108ijap0303_3.

[Ken05] Rusty Kennedy. "CMU puts words in Ben Franklin's mouth". In: *Pittsburgh Post-Gazette (June 30th)* (2005). URL: https://www.post-gazette.com/news/science/2005/06/30/CMU-puts-words-in-Ben-Franklin-s-mouth/stories/200506300463 (visited on 09/21/2023).

[KFZ20] Katharina Kühne, Martin H. Fischer, and Yuefang Zhou. "The Human Takes It All: Humanlike Synthesized Voices Are Perceived as Less Eerie and More Likable. Evidence From a Subjective Ratings Study". In: *Frontiers in Neurorobotics* 14 (2020). ISSN: 1662-5218. DOI: 10.3389/fnbot.2020.593732.

[Koe+20] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. "Racial disparities in automated speech recognition". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.14 (2020), pp. 7684–7689. DOI: 10.1073/pnas.1915768117. URL: https://www.pnas.org/doi/full/10.1073/pnas.1915768117.

[Kre18] Daniel Kreps. "Roy Orbison Hologram Tour Sets North American Dates". In: *Rolling Stone* (2018). URL: https://www.rollingstone.com/music/music-news/roy-orbison-hologram-tour-sets-north-american-dates-694945/ (visited on 03/27/2024).

[KWA72] Elisabeth Kübler-Ross, Stanford Wessler, and Louis V. Avioli. "On Death and Dying". In: *JAMA* 221.2 (1972), pp. 174–179. DOI: 10.1001/jama.1972.03200150040010.

[KY18] Sarah Kenderdine and Andrew Yip. "The proliferation of aura: Facsimiles, authenticity and digital objects". In: *The Routledge Handbook of Museums, Media and Communication*. Ed. by Kirsten Drotner, Vince Dziekan, Ross Parry, and Kim Christian Schrøder. London: Routledge, 2018.

[Lan22] Luca Lang. "Virtual-Reality-Projekt gegen Holocaust-Vergessen bei Jugend". In: *Süddeutsche Zeitung* (2022). URL: https://www.sueddeutsche.de/muenchen/muenchen-holocaust-virtual-reality-1.5619114 (visited on 08/28/2024).

[Lee06] Seong-Whan Lee. "Automatic Gesture Recognition for Intelligent Human-Robot Interaction". In: *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. IEEE, 2006. DOI: 10.1109/fgr.2006.25.

[Liu+18] Can Liu, Ben Bengler, Danilo Di Cuia, Katie Seaborn, Giovanna Nunes Vilaza, Sarah Gallacher, Licia Capra, and Yvonne Rogers. "Pinsight". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. New York, NY, USA: ACM, 2018. DOI: 10.1145/3196709.3196782.

[Liu+23] Yimeng Liu, Jacob Ritchie, Sven Kratz, Misha Sra, Brian A. Smith, Andrés Monroy-Hernández, and Rajan Vaish. "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*. New York, NY, USA: ACM, 2023. DOI: 10.1145/3544549.3585588.

[LL11] Bruno Latour and Adam Lowe. "The migration of the aura, or how to explore the original through its facsimiles". In: *Switching codes*. Ed. by Thomas Bartscherer and Roderick Coover. Chicago, Ill and London: University of Chicago Press, 2011, pp. 275–298. ISBN: 9780226038322. DOI: 10.7208/9780226038322-017.

[Lou+17] Emilie Loup-Escande, Eric Jamet, Martin Ragot, Séverine Erhel, and Nicolas Michinov. "Effects of Stereoscopic Display on Learning and User Experience in an Educational Virtual Environment". In: *International Journal of Human–Computer Interaction* 33.2 (2017), pp. 115–122. DOI: 10.1080/10447318.2016.1220105.

[Lov+20] Kate Loveys, Gabrielle Sebaratnam, Mark Sagar, and Elizabeth Broadbent. "The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review". In: *International Journal of Social Robotics* 12.6 (2020), pp. 1293–1312. ISSN: 1875-4805. DOI: 10.1007/s12369-020-00680-7. URL: https://link.springer.com/article/10.1007/s12369-020-00680-7.

[Lov43] Ada Lovelace. "Translator's Notes to M. Menabrea's Memoir". In: *Scientific Memoirs, Selected from the Transactions of Foreign Academies of Science and Learned Societies and from Foreign Journals* 3 (1843). Ed. by Richard Taylor, pp. 691–731.

[Mak24] Mykola Makhortykh. *AI and the Holocaust: rewriting history? The impact of artificial intelligence on understanding the Holocaust.* Ed. by Heather Mann. 2024. DOI: 10.54675/ZHJC6844.

[Mal+15] 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. "EEG based evaluation of stereoscopic 3D displays for viewer discomfort". In: *BioMedical Engineering OnLine* 14.1 (2015), p. 21. ISSN: 1475-925X. DOI: 10.1186/s12938-015-0006-8.

[Mar+17] Jorge Martín-Gutiérrez, Carlos Efrén Mora, Beatriz Añorbe-Díaz, and Antonio González-Marrero. "Virtual Technologies Trends in Education". In: *Eurasia Journal of Mathematics, Science and Technology Education* 13.2 (2017), pp. 469–486. ISSN: 1305-8223. DOI: 10.12973/eurasia.2017.00626a. URL: https://www.ejmste.com/article/virtual-technologies-trends-in-education-4674.

[Mar+22] Alan S. Marcus, Rotem Maor, Ian M. McGregor, Gary Mills, Simone Schweber, Jeremy Stoddard, and David Hicks. "Holocaust education in transition from live to virtual survivor testimony: pedagogical and ethical dilemmas". In: *Holocaust Studies* 28.3 (2022), pp. 279–301. DOI: 10.1080/17504902.2021.1979176.

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

[May+23] Elisabeth Mayer, Rubén Jesús García Hernández, Daniel Kolb, Jutta Dreer, Simone Müller, Thomas Odaker, and Dieter Kranzlmüller. "Ten years Center for Immersive Visualizations - Past, Present, and Future". In: *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2023. DOI: 10.1109/vrw58643.2023.00051.

[May14] Richard E. Mayer. "Principles Based on Social Cues in Multimedia Learning: Personalization, Voice, Image, and Embodiment Principles". In: *The Cambridge Handbook of Multimedia Learning*. Ed. by Richard E. Mayer. Cambridge handbooks in psychology. New York, NY: Cambridge University Press, 2014, pp. 345–368. ISBN: 9781107035201. DOI: 10.1017/CBO9781139547369.017.

[MB11] Michael Massimi and Ronald M. Baecker. "Dealing with death in design". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2011. DOI: 10.1145/1978942.1979092.

[McM+19] Donald McMillan, Barry Brown, Ikkaku Kawaguchi, Razan Jaber, Jordi Solsona Belenguer, and Hideaki Kuzuoka. "Designing with Gaze". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–26. DOI: 10.1145/3359278.

[McM16] Thomas McMullan. *The virtual Holocaust survivor: how history gained new dimensions*. 2016. URL: https://www.theguardian.com/technology/2016/jun/18/holocaust-survivor-hologram-pinchas-gutter-new-dimensions-history (visited on 02/22/2024).

[MCW15] Minhua Ma, Sarah Coward, and Chris Walker. "Interact: a mixed reality virtual survivor for holocaust testimonies". In: *OzCHI '15: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. New York, NY, USA: Association for Computing Machinery, 2015, pp. 250–254. ISBN: 9781450336734. DOI: 10.1145/2838739.2838803.

[MCW17] Minhua Ma, Sarah Coward, and Chris Walker. "Question-Answering Virtual Humans Based on Pre-recorded Testimonies for Holocaust Education". In: *Serious Games and Edutainment Applications*. Springer, Cham, 2017, pp. 391–409. DOI: 10.1007/978-3-319-51645-5_18.

[MDK18] Scott E. Maxwell, Harold D. Delaney, and Ken Kelley. *Designing experiments and analyzing data: A model comparison perspective*. Third edition. New York and London: Routledge, 2018. ISBN: 9781315642956. DOI: 10.4324/9781315642956. URL: https://www.taylorfrancis.com/books/mono/10.4324/9781315642956/designing-experiments-analyzing-data-scott-maxwell-harold-delaney-ken-kelley.

[ME12] Christos N. Moridis and Anastasios A. Economides. "Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions". In: *IEEE Transactions on Affective Computing* 3.3 (2012), pp. 260–272. ISSN: 1949-3045. DOI: 10.1109/t-affc.2012.6.

[MI06] Karl F. MacDorman and Hiroshi Ishiguro. "The uncanny advantage of using androids in cognitive and social science research". In: *Interaction Studies* 7.3 (2006), pp. 297–337. ISSN: 1572-0381. DOI: 10.1075/is.7.3.03mac. URL: https://www.jbe-platform.com/content/journals/10.1075/is.7.3.03mac.

[Mit+11] Wade J. Mitchell, Kevin A. Szerszen, Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. MacDorman. "A mismatch in the human realism of face and voice produces an uncanny valley". In: *i-Perception* 2.1 (2011), pp. 10–12. DOI: 10.1068/i0415.

[Moo17a] Roger K. Moore. "Appropriate voices for artefacts: Some key insights". In: *Proceedings of the 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots*. Ed. by Angela Dassow, Ricard Marxer, and Roger K. Moore. 2017, pp. 7–11. URL: https://vihar-2017.vihar.org/assets/vihar2017_proceedings.pdf.

[Moo17b] Roger K. Moore. "Is Spoken Language All-or-Nothing? Implications for Future Speech-Based Human-Machine Interaction". In: *Dialogues with social robots*. Ed. by Kristiina Jokinen and Graham Wilcock. Lecture notes in electrical engineering. Singapore: Springer, 2017, pp. 281–291. ISBN: 9789811025846. DOI: 10.1007/978-981-10-2585-3_22.

[Moo23] Pete Moore. *The Pain Toolkit*. 2023. URL: https://www.paintoolkit.org/ (visited on 02/27/2024).

[Mor70] Masahiro Mori. "Bukimi no tani genshō [the uncanny valley]". In: *Energy* 7.4 (1970), pp. 33–35.

[Mos12] Jana M. Moser. "Tupac Lives: What Hologram Authors Should Know about Intellectual Property Law". In: *Business Law Today* (2012), pp. 1–5. URL: https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/busiltom2012%C2%A7ion=77.

[MS98] Donald Marinelli and Scott Stevens. "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. New York, NY, USA: Association for Computing Machinery, 1998, pp. 11–16. ISBN: 1581131623. DOI: 10.1145/306774.306780.

[Muh15] Muhanna A. Muhanna. "Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions". In: *Journal of King Saud University - Computer and Information Sciences* 27.3 (2015), pp. 344–361. ISSN: 1319-1578. DOI: 10.1016/j.jksuci.2014.03.023.

[MWM19] Guido Makransky, Philip Wismer, and Richard E. Mayer. "A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation". In: *Journal of Computer Assisted Learning* 35.3 (2019), pp. 349–358. ISSN: 1365-2729. DOI: 10.1111/jcal.12335.

[Nat22] National Marine Mammal Foundation. *Dr. Sam Ridgway: Storyfile Experience*. 2022. URL: https://www.nmmf.org/sam-ridgway-storyfile-experience/ (visited on 02/27/2024).

[Nav+23] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. *A Comprehensive Overview of Large Language Models*. 2023. DOI: 10.48550/arXiv.2307.06435.

[NCN23] Anika Nissen, Colin Conrad, and Aaron Newman. “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*. New York, NY, USA: ACM, 2023. DOI: 10.1145/3544548.3580943.

[NG99] Clifford Nass and Li Gong. “Maximized modality or constrained consistency?” In: *International Conference on Auditory-Visual Speech Processing*. Ed. by D. W. Massaro. 1999. URL: https://www.isca-speech.org/archive_open/avsp99/av99_001.html.

[NST94] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. “Computers are social actors”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Ed. by Beth Adelson. New York, NY: ACM, 1994, pp. 72–78. ISBN: 0897916506. DOI: 10.1145/191666.191703.

[NZ14] Abba Naor and Helmut Zeller. *Ich sang für die SS: Mein Weg vom Ghetto zum israelischen Geheimdienst*. C.H.Beck, 2014. ISBN: 9783406659843.

[OBW18] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. “A Systematic Review of Social Presence: Definition, Antecedents, and Implications”. In: *Frontiers in Robotics and AI* 5 (2018). ISSN: 2296-9144. DOI: 10.3389/frobt.2018.00114. URL: https://www.frontiersin.org/articles/10.3389/frobt.2018.00114/full.

[Odo+09] William Odom, James Pierce, Erik Stolterman, and Eli Blevis. “Understanding why we preserve some things and discard others in the context of interaction design”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2009. DOI: 10.1145/1518701.1518862.

[Pap+19] George Papanastasiou, Athanasios Drigas, Charalabos Skianis, Miltiadis Lytras, and Effrosyni Papanastasiou. “Virtual and augmented reality effects on K-12, higher and tertiary education students’ twenty-first century skills”. In: *Virtual Reality* 23.4 (2019), pp. 425–436. ISSN: 1434-9957. DOI: 10.1007/s10055-018-0363-2. URL: https://link.springer.com/article/10.1007/s10055-018-0363-2.

[Par+22] Dhaval Parmar, Stefan Olafsson, Dina Utami, Prasanth Murali, and Timothy Bickmore. “Designing Empathic Virtual Agents: Manipulating Animation, Voice, Rendering, and Empathy to Create Persuasive Agents”. In: *Autonomous Agents and Multi-Agent Systems* 36.1 (2022), pp. 1–24. ISSN: 1573-7454. DOI: 10.1007/s10458-021-09539-1.

[Pat+22] Pat Pataranutaporn, Joanne Leong, Valdemar Danry, Alyssa P. Lawson, Pattie Maes, and Misha Sra. “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, 2022. DOI: 10.1109/fie56618.2022.9962478.

[Pat+23] Pat Pataranutaporn, Valdemar Danry, Lancelot Blanchard, Lavanay Thakral, Naoki Ohsugi, Pattie Maes, and Misha Sra. “Living Memories: AI-Generated Characters as Digital Mementos”. In: *Proceedings of the 28th International Conference on Intelligent User Interfaces*. New York, NY, USA: ACM, 2023. DOI: 10.1145/3581641.3584065.

[PCS21] Delphine Potdevin, Céline Clavel, and Nicolas Sabouret. “Virtual intimacy in human-embodied conversational agent interactions: the influence of multimodality on its perception”. In: *Journal on Multimodal User Interfaces* 15.1 (2021), pp. 25–43. ISSN: 1783-8738. DOI: 10.1007/s12193-020-00337-9. URL: https://link.springer.com/article/10.1007/s12193-020-00337-9.

[Pek+11] Reinhard Pekrun, Thomas Goetz, Anne C. Frenzel, Petra Barchfeld, and Raymond P. Perry. “Measuring emotions in students’ learning and performance: The Achievement Emotions Questionnaire (AEQ)”. In: *Contemporary Educational Psychology* 36.1 (2011), pp. 36–48. ISSN: 0361-476X. DOI: 10.1016/j.cedpsych.2010.10.002.

[Phi+23] Rachel Phinnemore, Mohi Reza, Blaine Lewis, Karthik Mahadevan, Bryan Wang, Michelle Annett, and Daniel Wigdor. “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*. New York, NY, USA: ACM, 2023. DOI: 10.1145/3544548.3581346.

[Pil11] Pilotfish. *Infogate*. 2011. URL: https://www.pilotfish.eu/case/infogate/ (visited on 02/27/2024).

[PMM21] Gustav Bøg Petersen, Aske Mottelson, and Guido Makransky. "Pedagogical Agents in Educational VR: An in the Wild Study". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445760.

[Pom+20] Patryk Pomykalski, Mikołaj P. Woźniak, Paweł W. Woźniak, Krzysztof Grudzień, Shengdong Zhao, and Andrzej Romanowski. "Considering Wake Gestures for Smart Assistant Use". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2020. DOI: 10.1145/3334480.3383089.

[Pop63] Karl R. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge, 1963.

[RA19] George E. Raptis and Nikolaos M. Avouris. "Supporting Designers in Creating Cognition-centered Personalized Cultural Heritage Activities". In: *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. Ed. by George Angelos Papadopoulos. ACM Digital Library. New York,NY,United States: Association for Computing Machinery, 2019, pp. 407–411. ISBN: 9781450367110. DOI: 10.1145/3314183.3323868.

[Raf88] Sheizf Rafaeli. "Interactivity: From new media to communication". In: *Sage annual review of communication research: Advancing communication science* 16 (1988), pp. 110–134.

[Ree19a] Ryan Reed. "Roy Orbison, Buddy Holly Holograms Paired for Fall Tour". In: *Rolling Stone* (2019). URL: https://www.rollingstone.com/music/music-news/roy-orbison-buddy-holly-hologram-tour-812886/ (visited on 03/27/2024).

[Ree19b] Ryan Reed. "Whitney Houston Hologram Tour Plots First 2020 Dates". In: *Rolling Stone* (2019). URL: https://www.rollingstone.com/music/music-news/whitney-houston-hologram-tour-2020-dates-image-885666/ (visited on 03/27/2024).

[Rei78] Jasia Reichardt. *Robots: Fact, fiction, and prediction*. London: Thames and Hudson, 1978.

[Rhe+21] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. "Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design". In: *International Journal of Human–Computer Interaction* 37.1 (2021), pp. 81–96. DOI: 10.1080/10447318.2020.1807710.

[Ric21] Alasdair Richardson. "Touching distance: young people's reflections on hearing testimony from Holocaust survivors". In: *Journal of Modern Jewish Studies* 20.3 (2021), pp. 315–338. DOI: 10.1080/14725886.2021.1874692.

[RM87] Rebecca B. Rubin and Michael P. McHugh. "Development of parasocial interaction relationships". In: *Journal of Broadcasting & Electronic Media* 31.3 (1987), pp. 279–292. ISSN: 0883-8151. DOI: 10.1080/08838158709386664.

[Ron22] Shlomi Ron. *What is Conversational Video?* 2022. URL: https://www.visualstorytell.com/blog/what-is-conversational-video (visited on 09/13/2023).

[RP19] Eimear Ryan and Claire Poole. "Impact of Virtual Learning Environment on Students' Satisfaction, Engagement, Recall, and Retention". In: *Journal of Medical Imaging and Radiation Sciences* 50.3 (2019), pp. 408–415. ISSN: 1939-8654. DOI: 10.1016/j.jmir.2019.04.005.

[RR94] Charles S. Reichardt and Sharon F. Rallis. "Qualitative and quantitative inquiries are not incompatible: A call for a new partnership". In: *New Directions for Program Evaluation* 1994.61 (1994), pp. 85–91. ISSN: 1551-2371. DOI: 10.1002/ev.1670.

[Rza+19] Rufat Rzayev, Gürkan Karaman, Katrin Wolf, Niels Henze, and Valentin Schwind. "The Effect of Presence and Appearance of Guides in Virtual Reality Exhibitions". In: *Proceedings of Mensch und Computer 2019*. New York, NY, USA: ACM, 2019, pp. 11–20. ISBN: 9781450371988. DOI: 10.1145/3340764.3340802.

[Sab12] Martin Sabrow. "Der Zeitzeuge als Wanderer zwischen zwei Welten". In: *Die Geburt des Zeitzeugen nach 1945*. Ed. by Martin Sabrow and Norbert Frei. Wallstein Verlag, 2012, pp. 11–32. ISBN: 978-3-8353-2245-5. DOI: 10.5771/9783835322455-11.

[Sab22] Martin Sabrow. "Zeitzeuge". In: *Handbuch Historische Authentizität*. Ed. by Martin Sabrow and Achim Saupe. Wallstein Verlag, 2022, pp. 553–562. ISBN: 978-3-8353-4795-3. DOI: 10.5771/9783835347953-553.

[Sch+22] Oliver Schreer, Markus Worchel, Rodrigo Diaz, Sylvain Renault, Wieland Morgenstern, Ingo Feldmann, Marcus Zepp, Anna Hilsmann, and Peter Eisert. "Preserving Memories of Contemporary Witnesses Using Volumetric Video". In: *i-com* 21.1 (2022), pp. 71–82. ISSN: 2196-6826. DOI: 10.1515/icom-2022-0015.

[Sch21] Corey Kai Nelson Schultz. "Creating the 'virtual' witness: the limits of empathy". In: *Museum Management and Curatorship* (2021), pp. 1–16. DOI: 10.1080/09647775.2021.1954980.

[Scu97] Raymond Scupin. "The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology". In: *Human Organization* 56.2 (1997), pp. 233–237. ISSN: 0018-7259. DOI: 10.17730/humo.56.2.x335923511444655.

[SD39] Walter Andrew Shewhart and William Edwards Deming. *Statistical method from the viewpoint of quality control*. 1st ed. Republication 1986. New York: Dover Publications, 1939. ISBN: 9780486652320.

[SEG17] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. "Intersectional HCI". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2017. DOI: 10.1145/3025453.3025766.

[SFR01] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. "The Experience of Presence: Factor Analytic Insights". In: *Presence: Teleoperators and Virtual Environments* 10.3 (2001), pp. 266–281. DOI: 10.1162/105474601300343603.

[SH21] Lisa Schwendemann and Ernst Hüttl. "1000 Fragen an Abba Naor und Eva Umlauf – Die Rolle der Fragen und Antworten für die Erstellung interaktiver 3D-Zeugnisse". In: *Interaktive 3D-Zeugnisse von Holocaust-Überlebenden. Chancen und Grenzen einer innovativen Technologie*. Ed. by Anja Ballis, Markus Gloe, Florian Duda, Fabian Heindl, Ernst Hüttl, Daniel Kolb, and Lisa Schwendemann. Vol. 1. Eckert. Dossiers. Braunschweig, Germany: Georg-Eckert-Institut - Leibniz-Institut für internationale Schulbuchforschung, 2021, pp. 25–42.

[She24] Harriet Sherwood. "Return of the king: Elvis hologram show to premiere in London". In: *The Guardian* (2024). URL: https://www.theguardian.com/music/2024/jan/04/elvis-evolution-hologram-show-london-premiere (visited on 03/27/2024).

[Sho+22] Michael Shorter, Bettina Minder, Jon Rogers, Matthias Baldauf, Aurelio Todisco, Sabine Junginger, Aysun Aytaç, and Patricia Wolf. "Materialising the Immaterial: Provotyping to Explore Voice Assistant Complexities". In: *Designing Interactive Systems Conference 2022*. Ed. by Florian Mueller. ACM Digital Library. New York, NY, United States: Association for Computing Machinery, 2022, pp. 1512–1524. ISBN: 9781450393584. DOI: 10.1145/3532106.3533519.

[SHT17] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. "Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S)". In: *International Journal of Interactive Multimedia and Artificial Intelligence* 4.6 (2017), pp. 103–108. ISSN: 1989-1660. DOI: 10.9781/ijimai.2017.09.001.

[SKF12] Cyrus Shahabi, Ali Khodaei, and Barak Fishbain. "Geo-Immersive Learning". In: *Pastforward* (2012).

[Sla09] Mel Slater. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364.1535 (2009), pp. 3549–3557. DOI: 10.1098/rstb.2009.0138.

[Sma21] Etan Smallman. "Freak out! It's Nile Rodgers in your living room, singing and answering questions". In: *The Guardian* (2021). URL: https://www.theguardian.com/artanddesign/2021/mar/24/freak-out-nile-rodgers-digital-portrait-living-room-singing-answering-questions (visited on 02/27/2024).

[SP20] Michal Shur-Ofry and Guy Pessach. "Robotic Collective Memory". In: *Washington University Law Review* 97.3 (2020), pp. 975–1005. DOI: 10.2139/ssrn.3364008.

[Spi+05] Lisa Spiro, Janice Bordeaux, Diane Butler, Andrea Martin, Chris Pound, Aniko Sandor, and Geneva Henry. "The Shoah Visual History Archive: Experience from the Classroom". In: *Proceedings of ED-Media 2005, World Conference on Educational Multimedia, Hypermedia & Telecommunications*. Ed. by Piet Kommers. Association for the Advancement of Computing in Education (AACE), 2005, pp. 355–360. ISBN: 978-1-880094-56-3. URL: https://www.learntechlib.org/p/20107/.

[Sto21] StoryFile. "William Shatner Joins New Frontier of AI at 90". In: (2021). Accessed: 2024-02-27. URL: https://www.globenewswire.com/news-release/2021/03/22/2196702/0/en/William-Shatner-Joins-New-Frontier-of-AI-at-90.html (visited on 02/27/2024).

[SU21] Sanna Stegmaier and Svetlana Ushakova. "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*. Ed. by Victoria Grace Walden. Cham: Springer, 2021, pp. 61–96. DOI: 10.1007/978-3-030-83496-8_4.

[Sun+12] S. Shyam Sundar, Jeeyun Oh, Saraswathi Bellur, Haiyan Jia, and Hyang-Sook Kim. "Interactivity as self-expression: A field experiment with customization and blogging". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2012. DOI: 10.1145/2207676.2207731.

[Sun+15] S. Shyam Sundar, Haiyan Jia, T. Franklin Waddell, and Yan Huang. "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*. Ed. by S. Shyam Sundar. Hoboken, New Jersey: John Wiley & Sons, Ltd, 2015, pp. 47–86. ISBN: 9781118426456. DOI: 10.1002/9781118426456.ch3.

[SWC76] John Short, Ederyn Williams, and Bruce Christie. *The social psychology of telecommunications*. London: Wiley, 1976. ISBN: 0471015814.

[TA16] Italo Trizano-Hermosilla and Jesús M. Alvarado. "Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements". In: *Frontiers in Psychology* 7 (2016), p. 769. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2016.00769.

[Tak23] Kanako Takahara. "Inside Hiroshima's Atomic Bomb Dome: Broadcaster taps tech to keep history alive". In: *The Japan Times* (2023). URL: https://www.japantimes.co.jp/news/2023/05/20/national/nhk-tech-hiroshima-history/ (visited on 02/27/2024).

[Tch+16] Katy Tcha-Tokey, Olivier Christmann, Emilie Loup-Escande, and Simon Richir. "Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments". In: *The International Journal of Virtual Reality* 16.1 (2016), pp. 33–48. URL: https://193.48.193.34/handle/10985/11352.

[TGN15] Angela Tinwell, Mark Grimshaw, and Deborah Abdel Nabi. "The effect of onset asynchrony in audio-visual speech and the Uncanny Valley in virtual characters". In: *International Journal of Mechanisms and Robotic Systems* 2.2 (2015), pp. 97–110. DOI: 10.1504/IJMRS.2015.068991.

[The23] The Matt Sky Podcast. *Sarah Coward, CEO, In The Room: Conversational Media, AI & Creating Meaningful Connections*. 2023. URL: https://www.youtube.com/watch?v=r196qxXKiL8 (visited on 01/23/2024).

[Tin+11] Angela Tinwell, Mark Grimshaw, Deborah Abdel Nabi, and Andrew Williams. "Facial expression of emotion and perception of the Uncanny Valley in virtual characters". In: *Computers in Human Behavior* 27.2 (2011), pp. 741–749. ISSN: 0747-5632. DOI: 10.1016/j.chb.2010.10.018.

[Tra+15a] David Traum, Kallirroi Georgila, Ron Artstein, and Anton Leuski. "Evaluating Spoken Dialogue Processing for Time-Offset Interaction". In: *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015. DOI: 10.18653/v1/w15-4629.

[Tra+15b] 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. "New Dimensions in Testimony: Digitally Preserving a Holocaust Survivor's Interactive Storytelling". In: *Interactive Storytelling*. Ed. by Henrik Schoenau-Fog, Luis Emilio Bruni, Sandy Louchart, and Sarune Baceviciute. Vol. 9445. Cham: Springer International Publishing, 2015, pp. 269–281. DOI: 10.1007/978-3-319-27036-4_26.

[Tra16] David Traum. "Using Dialogue System Technology to Support Interactive History Learning". In: *Journal of the Japanese Society for Artificial Intelligence* 31.6 (2016), pp. 806–812. ISSN: 2435-8614. DOI: 10.11517/jjsai.31.6_806_2.

[Tro17] Brenda Trofanenko. ""We Tell Stories": Oral History as a Pedagogical Encounter". In: *Oral History and Education*. Ed. by Kristina R. Llewellyn and Nicholas Ng-A-Fook. New York, NY, USA: Palgrave Macmillan, 2017, pp. 149–165. DOI: 10.1057/978-1-349-95019-5_8.

[Tur50] Alan Mathison Turing. "Computing Machinery and Intelligence". In: *Mind* LIX.236 (1950), pp. 433–460. ISSN: 0026-4423. DOI: 10.1093/mind/LIX.236.433.

[Uni21] United States Census Bureau. *2020 Population Distribution in the United States and Puerto Rico*. 2021. URL: https://www.census.gov/library/visualizations/2021/geo/population-distribution-2020.html (visited on 04/08/2024).

[UO16] Eva Umlauf and Stefanie Oswalt. *Die Nummer auf deinem Unterarm ist blau wie deine Augen: Erinnerungen*. Hoffmann und Campe, 2016.

[USC22] USC Shoah Foundation. *Dimensions in Testimony Reaches Milestone of 50 Interactive Interviews*. 2022. URL: https://sfi.usc.edu/news/2021/12/32271-dimensions-testimony-reaches-milestone-50-interactive-interviews (visited on 09/21/2023).

[VAP19] Mahesh Vanjani, Milam Aiken, and Mina Park. "Chatbots for multilingual conversations". In: *Journal of Management Science and Business Intelligence* 4.1 (2019), pp. 19–24. URL: http://ibii-us.org/journals/jmsbi/v4n1/publish/v4n1_3.pdf.

[Vas+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. *Attention Is All You Need*. 2017. DOI: 10.48550/arXiv.1706.03762.

[Ver+00] Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. "Why conversational agents should catch the eye". In: *CHI '00 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2000. DOI: 10.1145/633292.633442.

[Vir19] Graeme Virtue. "The Bizarre World of Frank Zappa review – toking poodles and holographic necromancy". In: *The Guardian* (2019). URL: https://www.theguardian.com/music/2019/may/10/the-bizarre-world-of-frank-zappa-review-edinburgh-playhouse (visited on 03/27/2024).

[Vis+22] Sruthi Viswanathan, Fabien Guillot, Minsuk Chang, Antonietta Maria Grasso, and Jean-Michel Renders. "Addressing Hiccups in Conversations with Recommender Systems". In: *Designing Interactive Systems Conference 2022*. Ed. by Florian Mueller. ACM Digital Library. New York,NY,United States: Association for Computing Machinery, 2022, pp. 1243–1259. ISBN: 9781450393584. DOI: 10.1145/3532106.3533491.

[vSS03] Janie A. van Dijk, Mirjam J. A. Schoutrop, and Philip Spinhoven. "Testimony therapy: treatment method for traumatized victims of organized violence". In: *American Journal of Psychotherapy* 57.3 (2003), pp. 361–373. DOI: 10.1176/appi.psychotherapy.2003.57.3.361.

[VSS05] Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. "Building characters: Lessons drawn from virtual environments". In: *Proceedings of toward social mechanisms of android science: A CogSci 2005 workshop*. 2005, pp. 119–126.

[VTM10] Valentijn T. Visch, Ed S. Tan, and Dylan Molenaar. "The emotional and cognitive effect of immersion in film viewing". In: *Cognition & Emotion* 24.8 (2010), pp. 1439–1445. ISSN: 0269-9931. DOI: 10.1080/02699930903498186.

[WCT88] David Watson, Lee Anna Clark, and Auke Tellegen. “Development and validation of brief measures of positive and negative affect: The PANAS scales”. In: *Journal of Personality and Social Psychology* 54.6 (1988), pp. 1063–1070. ISSN: 1939-1315. DOI: 10.1037/0022-3514.54.6.1063.

[Wie06] Annette Wieviorka. *The era of the witness*. Ithaca and London: Cornell University Press, 2006. ISBN: 9780801473166.

[WJ77] Brian Woodhouse and Paul H. Jackson. “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”. In: *Psychometrika* 42.4 (1977), pp. 579–591. ISSN: 1860-0980. DOI: 10.1007/BF02295980. URL: https://link.springer.com/article/10.1007/bf02295980.

[WM23] Victoria Grace Walden and Kate Marrison. *Recommendations for using Artificial Intelligence and Machine Learning for Holocaust Memory and Education*. 2023. DOI: 10.20919/elvh8804.

[WS98] Bob G. Witmer and Michael J. Singer. “Measuring Presence in Virtual Environments: A Presence Questionnaire”. In: *Presence: Teleoperators and Virtual Environments* 7.3 (1998), pp. 225–240. DOI: 10.1162/105474698565686.

[Xyg+23] Anna Xygkou, Panote Siriaraya, Alexandra Covaci, Holly Gwen Prigerson, Robert Neimeyer, Chee Siang Ang, and Wan-Jou She. “The "Conversation" about Loss: Understanding How Chatbot Technology was Used in Supporting People in Grief”. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Ed. by Albrecht Schmidt. ACM Digital Library. New York, NY, United States: Association for Computing Machinery, 2023, pp. 1–15. ISBN: 9781450394215. DOI: 10.1145/3544548.3581154.

[Yan+12] Shun-nan Yang, Tawny Schlieski, Brent Selmins, Scott C. Cooper, Rina A. Doherty, Philip J. Corriveau, and James E. Sheedy. “Stereoscopic viewing and reported perceived immersion and symptoms”. In: *Optometry and Vision Science* 89.7 (2012), pp. 1068–1080. ISSN: 1538-9235. DOI: 10.1097/OPX.0b013e31825da430. URL: https://journals.lww.com/optvissci/fulltext/2012/07000/stereoscopic_viewing_and_reported_perceived.17.aspx.

[Yoo+16] Daisy Yoo, Katie Derthick, Shaghayegh Ghassemian, Jean Hakizimana, Brian Gill, and Batya Friedman. “Multi-lifespan Design Thinking”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Ed. by Jofish Kaye, Allison Druin, Cliff Lampe, Dan Morris, and Juan Pablo Hourcade. ACM Digital Library. New York, NY: ACM, 2016, pp. 4423–4434. ISBN: 9781450333627. DOI: 10.1145/2858036.2858366.

[Yu+19] Qian Yu, Tonya Nguyen, Soravis Prakkamakul, and Niloufar Salehi. “"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*. New York, NY, USA: Association for Computing Machinery, 2019. ISBN: 978-1-4503-5971-9. DOI: 10.1145/3290607.3312918.

[Zei22] Steven Zeitchik. “At Walmart, Sam Walton’s hologram hints at Black Friday’s future”. In: *The Washington Post* (2022). URL: https://www.washingtonpost.com/technology/2022/11/25/walmart-black-friday-sam-walton/ (visited on 02/27/2024).

[Zie+20] Naim Zierau, Edona Elshan, Camillo Visini, and Andreas Janson. “A Review of the Empirical Literature on Conversational Agents and Future Research Directions”. In: *International Conference on Information Systems (ICIS)*. 2020. ISBN: 978-1-7336325-5-3. URL: https://www.alexandria.unisg.ch/261081/.

# Appendix

## A Supplementary Material to Expert Interviews

### A.1 Semi-structured Interview Guide

1. Introduction
   (a) Greeting, thank you for participating in this research session
   (b) Explain the topic and context
      i. I'm really interested in your perspective and experience
      ii. Everything you say will be kept confidential
   (c) Procedure
      i. Explain term "Interactive Digital Testimony"
   (d) Consent form
   (e) Any questions?
   (f) Start recording
2. Before we begin, would you mind telling me a bit about your involvement with interactive digital testimonies?
   (a) For how long?
   (b) What are/were your tasks?
3. What is the function/purpose of interactive digital testimonies?
   (a) Could you elaborate on that further?
   (b) Who is your target audience?
   (c) Have you encountered concerns or reservations regarding the use of interactive digital testimonies?
      i. E.g., Ethics, Legal, Technology
4. Which properties or features do interactive digital testimonies possess to achieve this function/purpose?
   (a) What is essential/important?
      i. Why is that?
   (b) What is optional?
      i. Why is that?
   (c) Which features depend on/influence others?
5. Could you walk me through a typical session (recording, production, use)?
   (a) Which technology/features do you use?
      i. Why?
   (b) What works well/as it should?

   (c) Have you encountered issues or problems of any kind? Tell me about a time when something didn't work out.
      i. Why did that happen?

6. Given unlimited resources and technological progress, what does the perfect interactive digital testimony look like? / What would you improve about interactive digital testimonies?
   (a) How would that improve interactive digital testimonies?
   (b) Why is that important to you?
   (c) What should not be added/changed?
      i. Why?

7. Summarize and outro
   (a) Main statements
   (b) Is there anything else you'd like to add?
   (c) Thank you
   (d) Stop Recording
   (e) Do you know of other people who might provide valuable insights?
   (f) Goodbye

## A.2 Interview Details

**Table A.1:** Detailed overview of interviewees and the circumstances of the respective interviews.

| ID | Duration | in hours | Language | Roles | Experience | Setting | Location |
|---|---|---|---|---|---|---|---|
| P1 | 01:24:42 | 1.41 | DE | Producer, Facilitator, Researcher | 3 years | In-person | Workplace, Office |
| P2 | 01:17:44 | 1.30 | DE | Contemporary Witness | 4 years | In-Person | Home |
| P3 | 00:31:53 | 0.53 | EN | Facilitator | 6 years | In-Person | Workplace, Museum |
| P4 | 00:48:01 | 0.80 | EN | Facilitator | 6 years | In-Person | Workplace, Museum |
| P5 | 01:09:28 | 1.16 | EN | Facilitator, Management | 5 years | In-Person | Workplace, Museum |
| P6 | 00:52:02 | 0.87 | EN | Management | 1 year | In-Person | Workplace, Office |
| P7 | 01:15:39 | 1.26 | EN | Facilitator | 1 year | In-Person | Workplace, Office |
| P8 | 02:22:38 | 2.38 | EN | Facilitator | 3 years | In-Person | Workplace, Office |
| P9 | 01:22:29 | 1.37 | EN | Facilitator | 3 years | In-Person | Workplace, Office |
| P10 | 01:51:05 | 1.85 | EN | Facilitator | <1 year | In-Person | Workplace, Office |
| P11 | 00:43:25 | 0.72 | EN | Facilitator, Management | 3 years | In-Person | Workplace, Office |
| P12 | 00:42:01 | 0.70 | EN | Management | 5 years | In-Person | Workplace, Office |
| P13 | 00:28:02 | 0.47 | EN | Management | 5 years | In-Person | Workplace, Office |
| P14 | 00:39:52 | 0.66 | DE | Contemporary Witness | 1 year | In-Person | Home |
| P15 | 01:02:45 | 1.05 | EN | Researcher | 5 years | In-Person | Workplace, Office |
| P16 | 01:27:06 | 1.45 | EN | Contemporary Witness | 7 years | Video call | Home |
| P17 | 00:58:20 | 0.97 | EN | Bereaved | 6 years | Video call | Home |
| P18 | 01:23:15 | 1.39 | EN | Management | <1 year (5 years before interview) | Video call | Home |
| P19 | 00:55:43 | 0.93 | EN | Management | <1 year (5 years before interview) | Video call | Home |
| P20 | 01:37:24 | 1.62 | EN | Producer | 13 years | Video call | Office |
| P21 | 02:13:43 | 2.23 | EN | Producer, Bereaved, Researcher | 13 years | Video call | Home |
| P22 | 00:52:00 | 0.87 | DE | Producer, Facilitator, Researcher | 4 years | In-Person | Workplace, Office |
| P23 | 01:06:37 | 1.11 | EN | Producer | 9 years | Video call | Home |
| P24 | 02:00:06 | 2.00 | DE | Producer, Facilitator, Researcher | 3 years | Video call | Home |
| P25 | 01:42:16 | 1.70 | DE | Producer, Management, Researcher | 4 years | In-Person | Workplace, Office |
| P26 | 01:51:16 | 1.85 | EN | Contemporary Witness | 2 years | In-Person | Interviewer's own office |

# B Supplementary Material to Study 1

## B.1 Questionnaire Items

**Table B.1:** Translated wording of our questionnaire items to measure User Experience and Presence in the audio-only vs. audio-visual 2D study. We provide the phrasing of the corresponding items in the original questionnaires as a comparison. Item WS-5 was only present in the questionnaire for participants who interacted with the audio-visual 2D IDT.

| ID | Item | Equivalent in original questionnaire |
|---|---|---|
| UEQ-S-2 | Asking questions was easy. | The application was easy to use. |
| UEQ-S-3 | The digital Holocaust survivor was able to answer my questions. | The application was efficient. |
| UEQ-S-5 | I found the way the interactive digital testimony was presented exciting. | The application was exciting. |
| UEQ-S-6 | I found the conversation interesting. | The application was interesting. |
| IPQ-SP3 | I did not feel like I was in a virtual conversation. | I did not feel present in the virtual space. |
| IPQ-INV2 | I was not aware of my real environment. | I was not aware of my real environment. |
| IPQ-INV4 | I was completely captivated by the interactive digital testimony. | I was completely captivated by the virtual world. |
| WS-5 | The visual aspects of the interactive digital testimony engaged me. | How much did the visual aspects of the environment involve you? |
| WS-6 | The auditory aspects of the interactive digital testimony engaged me. | How much did the auditory aspects of the environment involve you? |
| WS-8 | I was aware of events occurring in the real world around me. | How aware were you of events occurring in the real world around you? |
| WS-9 | I was not aware of the display and input devices. | How aware were you of your display and control devices? |
| WS-21 | It was difficult to interact with the interactive digital testimony. | How well could you move or manipulate objects in the virtual environment? |
| WS-23 | I asked the interactive digital testimony many questions. | How involved were you in the virtual environment experience? |
| WS-25 | The interactive digital testimony reacted quickly to my questions. | How much delay did you experience between your actions and expected outcomes? |
| WS-26 | I asked the interactive digital testimony questions right away. | How quickly did you adjust to the virtual environment experience? |
| WS-32 | The time passed quickly during the conversation. | Were you involved in the experimental task to the extent that you lost track of time? |

**Table B.2:** The PANAS items used in our questionnaire to measure users' emotions during the interaction. This table does not list correspondents in the original questionnaire since we did not adjust the wording of the items.

**(a)** Positive Emotions

| **ID** | **Item** |
|---|---|
| PANAS-01 | Active |
| PANAS-03 | Interested |
| PANAS-04 | Excited |
| PANAS-06 | Strong |
| PANAS-10 | Inspired |
| PANAS-11 | Proud |
| PANAS-13 | Enthusiastic |
| PANAS-15 | Alert |
| PANAS-17 | Determined |
| PANAS-18 | Attentive |

**(b)** Negative Emotions

| **ID** | **Item** |
|---|---|
| PANAS-02 | Distressed |
| PANAS-05 | Upset |
| PANAS-07 | Guilty |
| PANAS-08 | Scared |
| PANAS-09 | Hostile |
| PANAS-12 | Irritable |
| PANAS-14 | Ashamed |
| PANAS-16 | Nervous |
| PANAS-19 | Jittery |
| PANAS-20 | Afraid |

**Table B.3:** Translated wording of our questionnaire items to measure Accessibility.

| **ID** | **Item** | **Corresponding Guideline** |
|---|---|---|
| WCAG-1 | Notwithstanding sensory impairment, every person can perceive the interactive digital testimony well. | Information and user interface components must be presentable to users in ways they can perceive. |
| WCAG-2 | The interactive digital testimony is operable without restrictions. | User interface components and navigation must be operable. |
| WCAG-3 | Using the interactive digital testimony is easy to understand. | Information and the operation of the user interface must be understandable. |
| WCAG-4 | This type of user interaction can be performed by everyone. | Content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies. |

**Table B.4:** We varied the phrasing of our question about the preferred output modality, depending on the IDT modality the participants experienced during our study.

| **ID** | **Item for audio-only group** | **Item for audio-visual 2D group** |
|---|---|---|
| Preference | I would have preferred the accompanying video of the contemporary witness. | I would have preferred the voice-only presentation of the contemporary witness. |

## B.2 Interview Items

**Table B.5:** Questions asked during the fully structured interview.

| ID | Item |
|---|---|
| 1 | How did you like the method of presentation? |
| 2 | Was the interaction with the interactive digital testimony similar to a real conversation with a Holocaust survivor? |
| 3 | What did you feel during the interaction? |
| 4 | Do you feel the display type is suitable for use in educational institutions such as schools, universities, or museums? Where would you use this type of interaction? |
| 5 | Can the interaction be perceived by all people, including people with sensory impairments such as hearing loss or low vision? |
| 6 | If not, what is the reason that not everyone can participate and how can this be remedied? |
| 7 | Is there anything else that you would like to add? |

## B.3 Results of Statistical Analysis

Since item WS-5 was gathered only from participants who interacted with the audio-visual 2D IDT, it was omitted from individual comparative analysis. We reverse-keyed items IPQ-SP3, IPQ INV2, WS-8, and WS-21 during our analysis.

### B.3.1 Comparison between Groups with Different Output Modalities

**Table B.6:** Full data for comparison of modality. Mean, standard deviation, 95% confidence interval. p significance, F statistic (1 degree of freedom, Residuals 78), eta squared is effect size. Audio is audio-only, AV2D is audio-visual 2D.

| Variable | Audio | | AV2D | | p | F | $\eta^2$ |
|---|---|---|---|---|---|---|---|
| | M (SD) | $CI_{95\%}$ | M (SD) | $CI_{95\%}$ | | | |
| **User Exp.** | 3.74 (0.53) | ±0.18 | 3.97 (0.66) | ±0.20 | 0.09 | 3.02 | 0.04 |
| UEQ-S-2 | 4.17 (1.03) | ±0.35 | 4.33 (0.87) | ±0.26 | 0.41 | 0.68 | 5.7E-03 |
| UEQ-S-3 | 3.22 (0.83) | ±0.28 | 3.30 (1.11) | ±0.33 | 0.71 | 0.14 | 2.0E-03 |
| UEQ-S-5 | 3.72 (1.06) | ±0.36 | 4.59 (0.62) | ±0.18 | 1.3E-05 | 21.64 | 0.21 |
| UEQ-S-6 | 3.86 (0.87) | ±0.29 | 3.67 (1.25) | ±0.37 | 0.44 | 0.60 | 6.6E-03 |
| **Presence** | 3.08 (0.49) | ±0.16 | 3.23 (0.44) | ±0.13 | 0.15 | 2.13 | 0.03 |
| IPQ-SP3 | 3.53 (1.13) | ±0.38 | 3.24 (1.14) | ±0.34 | 0.26 | 1.30 | 0.02 |
| IPQ-INV2 | 2.17 (1.08) | ±0.37 | 2.15 (1.01) | ±0.30 | 0.95 | 4.0E-03 | 2.4E-06 |
| IPQ-INV4 | 3.06 (0.95) | ±0.32 | 3.37 (1.08) | ±0.32 | 0.17 | 1.88 | 0.02 |
| WS-6 | 4.33 (0.79) | ±0.27 | 4.37 (0.77) | ±0.23 | 0.84 | 0.04 | 4.0E-04 |
| WS-8 | 2.39 (1.05) | ±0.36 | 2.39 (1.04) | ±0.31 | 0.99 | 1.1E-04 | 2.3E-05 |
| WS-9 | 2.22 (1.12) | ±0.38 | 2.02 (1.02) | ±0.30 | 0.41 | 0.70 | 9.0E-03 |
| WS-21 | 2.72 (1.19) | ±0.40 | 3.00 (1.14) | ±0.34 | 0.28 | 1.16 | 0.01 |
| WS-23 | 3.31 (0.95) | ±0.32 | 3.11 (1.04) | ±0.31 | 0.37 | 0.83 | 0.01 |
| WS-25 | 3.58 (1.00) | ±0.34 | 4.07 (0.93) | ±0.28 | 0.02 | 5.72 | 0.06 |
| WS-26 | 2.83 (0.97) | ±0.33 | 2.83 (1.00) | ±0.30 | 0.97 | 1.1E-03 | 2.5E-05 |
| WS-32 | 3.69 (0.86) | ±0.29 | 3.85 (1.05) | ±0.31 | 0.48 | 0.49 | 6.5E-03 |
| **Pos. Emotions** | 2.75 (0.47) | ±0.16 | 2.76 (0.59) | ±0.17 | 0.98 | 9.6E-04 | 1.8E-06 |
| PANAS-01 | 3.19 (0.98) | ±0.33 | 3.04 (0.94) | ±0.28 | 0.49 | 0.49 | 6.0E-03 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| PANAS-03 | 4.42 (0.55) | ±0.19 | 4.26 (0.65) | ±0.19 | 0.25 | 1.34 | 0.02 |
| PANAS-04 | 1.64 (0.87) | ±0.29 | 1.85 (1.11) | ±0.33 | 0.35 | 0.88 | 9.6E-03 |
| PANAS-06 | 1.67 (1.01) | ±0.34 | 1.59 (0.93) | ±0.28 | 0.72 | 0.13 | 1.5E-03 |
| PANAS-10 | 2.64 (1.17) | ±0.40 | 2.76 (1.14) | ±0.34 | 0.64 | 0.22 | 2.5E-03 |
| PANAS-11 | 1.19 (0.62) | ±0.21 | 1.41 (0.86) | ±0.25 | 0.21 | 1.63 | 0.02 |
| PANAS-13 | 2.28 (1.43) | ±0.48 | 2.67 (1.28) | ±0.38 | 0.19 | 1.77 | 0.02 |
| PANAS-15 | 3.78 (1.07) | ±0.36 | 3.43 (1.00) | ±0.30 | 0.14 | 2.19 | 0.03 |
| PANAS-17 | 2.47 (1.30) | ±0.44 | 2.24 (0.97) | ±0.29 | 0.36 | 0.86 | 0.01 |
| PANAS-18 | 4.25 (0.84) | ±0.28 | 4.30 (0.79) | ±0.23 | 0.76 | 0.09 | 1.1E-03 |
| **Neg. Emotions** | 2.75 (0.47) | ±0.16 | 2.76 (0.59) | ±0.17 | 0.98 | 9.6E-04 | 1.8E-06 |
| PANAS-02 | 3.08 (1.25) | ±0.42 | 2.83 (1.23) | ±0.37 | 0.33 | 0.95 | 9.1E-03 |
| PANAS-05 | 1.75 (1.25) | ±0.42 | 1.46 (0.86) | ±0.26 | 0.21 | 1.58 | 0.02 |
| PANAS-07 | 1.78 (1.02) | ±0.34 | 1.59 (0.83) | ±0.25 | 0.36 | 0.86 | 0.01 |
| PANAS-08 | 3.17 (1.30) | ±0.44 | 2.70 (1.17) | ±0.35 | 0.09 | 3.02 | 0.04 |
| PANAS-09 | 1.08 (0.37) | ±0.12 | 1.15 (0.47) | ±0.14 | 0.48 | 0.51 | 6.2E-03 |
| PANAS-12 | 1.28 (0.74) | ±0.25 | 1.24 (0.60) | ±0.18 | 0.79 | 0.07 | 6.7E-04 |
| PANAS-14 | 2.47 (1.36) | ±0.46 | 2.15 (1.40) | ±0.42 | 0.31 | 1.07 | 0.01 |
| PANAS-16 | 1.89 (1.12) | ±0.38 | 1.98 (1.13) | ±0.33 | 0.72 | 0.13 | 1.7E-03 |
| PANAS-19 | 1.86 (1.20) | ±0.41 | 1.85 (0.87) | ±0.26 | 0.95 | 3.3E-03 | 5.0E-05 |
| PANAS-20 | 1.67 (1.10) | ±0.37 | 1.43 (0.72) | ±0.21 | 0.25 | 1.36 | 0.02 |
| **Accessibility** | 3.92 (0.62) | ±0.21 | 4.02 (0.58) | ±0.17 | 0.42 | 0.65 | 8.6E-03 |
| WCAG-1 | 3.69 (0.95) | ±0.32 | 3.63 (0.95) | ±0.28 | 0.76 | 0.09 | 7.7E-04 |
| WCAG-2 | 4.14 (0.80) | ±0.27 | 4.33 (0.70) | ±0.21 | 0.25 | 1.35 | 0.02 |
| WCAG-3 | 4.19 (0.92) | ±0.31 | 4.41 (0.80) | ±0.24 | 0.23 | 1.46 | 0.02 |
| WCAG-4 | 3.64 (1.17) | ±0.40 | 3.72 (1.03) | ±0.30 | 0.75 | 0.10 | 1.5E-03 |

### B.3.2 Comparison between Groups with different Participation Settings

**Table B.7:** Full data for comparison of setting. Mean, standard deviation, 95% confidence interval. p significance, F statistic (1 degree of freedom, Residuals 78), eta squared is effect size. Audio is audio-only, AV2D is audio-visual 2D.

| Variable | In-Person | | Online | | p | F | $\eta^2$ |
|---|---|---|---|---|---|---|---|
| | M (SD) | $CI_{95\%}$ | M (SD) | $CI_{95\%}$ | | | |
| **User Exp.** | 3.84 (0.55) | ±0.18 | 3.9 (0.67) | ±0.21 | 0.64 | 0.22 | 2.5E-03 |
| UEQ-S-2 | 4.58 (0.50) | ±0.16 | 3.95 (1.15) | ±0.36 | 1.9E-03 | 10.35 | 0.11 |
| UEQ-S-3 | 3.13 (0.99) | ±0.32 | 3.40 (0.99) | ±0.31 | 0.20 | 1.68 | 0.02 |
| UEQ-S-5 | 4.08 (0.94) | ±0.30 | 4.33 (0.93) | ±0.29 | 0.13 | 2.34 | 0.02 |
| UEQ-S-6 | 3.60 (1.03) | ±0.33 | 3.90 (1.14) | ±0.36 | 0.22 | 1.55 | 0.02 |
| **Presence** | 3.15 (0.43) | ±0.14 | 3.17 (0.50) | ±0.15 | 0.86 | 0.03 | 3.7E-04 |
| IPQ-SP3 | 3.53 (1.06) | ±0.34 | 3.21 (1.2) | ±0.37 | 0.21 | 1.60 | 0.02 |
| IPQ-INV2 | 1.95 (0.88) | ±0.28 | 2.36 (1.14) | ±0.36 | 0.08 | 3.18 | 0.04 |
| IPQ-INV4 | 3.25 (1.06) | ±0.34 | 3.21 (1.02) | ±0.32 | 0.91 | 0.01 | 1.7E-04 |
| WS-6 | 4.45 (0.64) | ±0.20 | 4.26 (0.89) | ±0.28 | 0.28 | 1.18 | 0.01 |
| WS-8 | 2.25 (0.98) | ±0.31 | 2.52 (1.09) | ±0.34 | 0.24 | 1.41 | 0.02 |
| WS-9 | 2.15 (1.17) | ±0.37 | 2.07 (0.97) | ±0.30 | 0.73 | 0.12 | 1.6E-03 |
| WS-21 | 3.00 (1.04) | ±0.33 | 2.76 (1.27) | ±0.39 | 0.37 | 0.81 | 0.01 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| WS-23 | 3.28 (1.09) | ±0.35 | 3.12 (0.92) | ±0.29 | 0.46 | 0.56 | 0.01 |
| WS-25 | 3.53 (0.91) | ±0.29 | 4.17 (0.96) | ±0.30 | 1.6E-03 | 10.72 | 0.11 |
| WS-26 | 2.88 (0.97) | ±0.31 | 2.79 (1.00) | ±0.31 | 0.68 | 0.17 | 2.1E-03 |
| WS-32 | 3.73 (0.99) | ±0.32 | 3.83 (0.96) | ±0.30 | 0.61 | 0.27 | 3.4E-03 |
| **Pos. Emotions** | 2.80 (0.59) | ±0.19 | 2.71 (0.49) | ±0.15 | 0.49 | 0.48 | 0.01 |
| PANAS-01 | 3.05 (1.01) | ±0.32 | 3.17 (0.91) | ±0.28 | 0.60 | 0.28 | 3.5E-03 |
| PANAS-03 | 4.30 (0.56) | ±0.18 | 4.36 (0.66) | ±0.20 | 0.69 | 0.16 | 1.9E-03 |
| PANAS-04 | 1.95 (1.08) | ±0.35 | 1.57 (0.91) | ±0.28 | 0.10 | 2.83 | 0.03 |
| PANAS-06 | 1.55 (1.01) | ±0.32 | 1.69 (0.92) | ±0.29 | 0.52 | 0.41 | 0.01 |
| PANAS-10 | 2.83 (1.24) | ±0.40 | 2.60 (1.06) | ±0.33 | 0.38 | 0.78 | 0.01 |
| PANAS-11 | 1.25 (0.71) | ±0.23 | 1.38 (0.82) | ±0.26 | 0.42 | 0.65 | 0.01 |
| PANAS-13 | 2.68 (1.40) | ±0.45 | 2.33 (1.30) | ±0.41 | 0.27 | 1.26 | 0.02 |
| PANAS-15 | 3.65 (1.03) | ±0.33 | 3.52 (1.06) | ±0.33 | 0.56 | 0.35 | 4.3E-03 |
| PANAS-17 | 2.43 (1.11) | ±0.35 | 2.26 (1.15) | ±0.36 | 0.50 | 0.46 | 0.01 |
| PANAS-18 | 4.30 (0.72) | ±0.23 | 4.26 (0.89) | ±0.28 | 0.84 | 0.04 | 5.2E-04 |
| **Neg. Emotions** | 2.80 (0.59) | ±0.19 | 2.71 (0.49) | ±0.15 | 0.49 | 0.48 | 0.01 |
| PANAS-02 | 2.58 (1.30) | ±0.42 | 3.29 (1.09) | ±0.34 | 9.0E-03 | 7.19 | 0.08 |
| PANAS-05 | 1.43 (0.87) | ±0.28 | 1.74 (1.19) | ±0.37 | 0.19 | 1.73 | 0.02 |
| PANAS-07 | 1.60 (0.84) | ±0.27 | 1.74 (0.99) | ±0.31 | 0.52 | 0.42 | 0.01 |
| PANAS-08 | 2.90 (1.26) | ±0.40 | 2.90 (1.25) | ±0.39 | 0.98 | 9.2E-04 | 1.1E-05 |
| PANAS-09 | 1.15 (0.48) | ±0.15 | 1.10 (0.37) | ±0.12 | 0.58 | 0.31 | 3.9E-03 |
| PANAS-12 | 1.18 (0.55) | ±0.18 | 1.33 (0.75) | ±0.24 | 0.28 | 1.16 | 0.01 |
| PANAS-14 | 2.25 (1.39) | ±0.44 | 2.33 (1.39) | ±0.43 | 0.81 | 0.06 | 7.4E-04 |
| PANAS-16 | 1.90 (1.08) | ±0.35 | 1.98 (1.16) | ±0.36 | 0.75 | 0.10 | 1.3E-03 |
| PANAS-19 | 1.88 (1.02) | ±0.33 | 1.83 (1.03) | ±0.32 | 0.86 | 0.03 | 4.3E-04 |
| PANAS-20 | 1.40 (0.87) | ±0.28 | 1.67 (0.93) | ±0.29 | 0.19 | 1.74 | 0.02 |
| **Accessibility** | 3.88 (0.60) | ±0.19 | 4.07 (0.58) | ±0.18 | 0.13 | 2.39 | 0.03 |
| WCAG-1 | 3.45 (1.04) | ±0.33 | 3.86 (0.81) | ±0.25 | 0.05 | 3.82 | 0.05 |
| WCAG-2 | 4.28 (0.68) | ±0.22 | 4.21 (0.81) | ±0.25 | 0.73 | 0.12 | 1.4E-03 |
| WCAG-3 | 4.18 (0.96) | ±0.31 | 4.45 (0.74) | ±0.23 | 0.12 | 2.48 | 0.03 |
| WCAG-4 | 3.60 (1.08) | ±0.35 | 3.76 (1.10) | ±0.34 | 0.50 | 0.45 | 0.01 |

### B.3.3 Comparison between Groups with different Output Modalities and Participation Settings

**Table B.8:** Full data for comparison of modality × setting. Mean, standard deviation, 95% confidence interval. p significance, F statistic (1 degree of freedom, Residuals 78), eta squared is effect size. Audio is audio-only, AV2D is audio-visual 2D.

| Variable | Audio * In-Person | | AV2D * In-Person | | Audio * Online | | AV2D * Online | | p | F | $\eta^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | M (SD) | $CI_{95\%}$ | M (SD) | $CI_{95\%}$ | M (SD) | $CI_{95\%}$ | M (SD) | $CI_{95\%}$ | | | |
| **User Exp.** | 3.90 (0.51) | ±0.26 | 3.80 (0.59) | ±0.26 | 3.61 (0.53) | ±0.26 | 4.14 (0.69) | ±0.30 | 0.02 | 5.65 | 0.06 |
| UEQ-S-2 | 4.76 (0.44) | ±0.22 | 4.43 (0.51) | ±0.22 | 3.63 (1.12) | ±0.54 | 4.22 (1.13) | ±0.49 | 0.02 | 5.58 | 0.06 |
| UEQ-S-3 | 3.29 (0.92) | ±0.47 | 3.00 (1.04) | ±0.45 | 3.16 (0.76) | ±0.37 | 3.61 (1.12) | ±0.48 | 0.09 | 2.89 | 0.03 |
| UEQ-S-5 | 3.65 (1.06) | ±0.54 | 4.39 (0.72) | ±0.31 | 3.79 (1.08) | ±0.52 | 4.78 (0.42) | ±0.18 | 0.51 | 0.45 | 4.4E-03 |
| UEQ-S-6 | 3.88 (0.78) | ±0.40 | 3.39 (1.16) | ±0.50 | 3.84 (0.96) | ±0.46 | 3.96 (1.30) | ±0.56 | 0.22 | 1.55 | 0.02 |
| **Presence** | 3.10 (0.45) | ±0.23 | 3.20 (0.42) | ±0.18 | 3.06 (0.53) | ±0.25 | 3.26 (0.47) | ±0.20 | 0.63 | 0.24 | 2.9E-03 |
| IPQ-SP3 | 3.71 (1.10) | ±0.57 | 3.39 (1.03) | ±0.45 | 3.37 (1.16) | ±0.56 | 3.09 (1.24) | ±0.54 | 0.95 | 4.3E-03 | 5.3E-05 |
| IPQ-INV2 | 1.88 (0.93) | ±0.48 | 2.00 (0.85) | ±0.37 | 2.42 (1.17) | ±0.56 | 2.30 (1.15) | ±0.50 | 0.61 | 0.26 | 3.2E-03 |
| IPQ-INV4 | 2.88 (0.93) | ±0.48 | 3.52 (1.08) | ±0.47 | 3.21 (0.98) | ±0.47 | 3.22 (1.09) | ±0.47 | 0.17 | 1.90 | 0.02 |
| WS-6 | 4.53 (0.62) | ±0.32 | 4.39 (0.66) | ±0.28 | 4.16 (0.9) | ±0.43 | 4.35 (0.88) | ±0.38 | 0.35 | 0.89 | 0.01 |
| WS-8 | 2.12 (0.99) | ±0.51 | 2.35 (0.98) | ±0.42 | 2.63 (1.07) | ±0.51 | 2.43 (1.12) | ±0.48 | 0.36 | 0.84 | 0.01 |
| WS-9 | 2.29 (1.16) | ±0.60 | 2.04 (1.19) | ±0.51 | 2.16 (1.12) | ±0.54 | 2.00 (0.85) | ±0.37 | 0.85 | 0.04 | 4.7E-04 |
| WS-21 | 3.00 (1.17) | ±0.60 | 3.00 (0.95) | ±0.41 | 2.47 (1.17) | ±0.57 | 3.00 (1.31) | ±0.57 | 0.31 | 1.04 | 0.01 |
| WS-23 | 3.71 (0.92) | ±0.47 | 2.96 (1.11) | ±0.48 | 2.95 (0.85) | ±0.41 | 3.26 (0.96) | ±0.42 | 0.02 | 6.01 | 0.07 |
| WS-25 | 3.35 (0.93) | ±0.48 | 3.65 (0.88) | ±0.38 | 3.79 (1.03) | ±0.50 | 4.48 (0.79) | ±0.34 | 0.34 | 0.93 | 0.01 |
| WS-26 | 3.00 (1.00) | ±0.51 | 2.78 (0.95) | ±0.41 | 2.68 (0.95) | ±0.46 | 2.87 (1.06) | ±0.46 | 0.36 | 0.83 | 0.01 |
| WS-32 | 3.59 (1.00) | ±0.52 | 3.83 (0.98) | ±0.43 | 3.79 (0.71) | ±0.34 | 3.87 (1.14) | ±0.49 | 0.72 | 0.13 | 1.6E-03 |
| **Pos. Emotions** | 2.85 (0.50) | ±0.26 | 2.76 (0.66) | ±0.28 | 2.67 (0.44) | ±0.21 | 2.75 (0.52) | ±0.23 | 0.49 | 0.49 | 0.01 |
| PANAS-01 | 3.18 (1.13) | ±0.58 | 2.96 (0.93) | ±0.40 | 3.21 (0.85) | ±0.41 | 3.13 (0.97) | ±0.42 | 0.75 | 0.11 | 1.3E-03 |
| PANAS-03 | 4.53 (0.51) | ±0.26 | 4.13 (0.55) | ±0.24 | 4.32 (0.58) | ±0.28 | 4.39 (0.72) | ±0.31 | 0.08 | 3.11 | 0.04 |
| PANAS-04 | 1.71 (0.85) | ±0.44 | 2.13 (1.22) | ±0.53 | 1.58 (0.90) | ±0.43 | 1.57 (0.95) | ±0.41 | 0.33 | 0.96 | 0.01 |
| PANAS-06 | 1.53 (1.07) | ±0.55 | 1.57 (0.99) | ±0.43 | 1.79 (0.98) | ±0.47 | 1.61 (0.89) | ±0.39 | 0.62 | 0.25 | 3.1E-03 |
| PANAS-10 | 2.82 (1.24) | ±0.64 | 2.83 (1.27) | ±0.55 | 2.47 (1.12) | ±0.54 | 2.70 (1.02) | ±0.44 | 0.67 | 0.18 | 2.3E-03 |

**Table B.8 continued from previous page**

| **Variable** | **Audio + In-Person** | | **AV2D + In-Person** | | **Audio + Online** | | **AV2D + Online** | | **p** | **F** | $\eta^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **M (SD)** | $\mathbf{CI_{95\%}}$ | **M (SD)** | $\mathbf{CI_{95\%}}$ | **M (SD)** | $\mathbf{CI_{95\%}}$ | **M (SD)** | $\mathbf{CI_{95\%}}$ | | | |
| PANAS-11 | 1.06 (0.24) | ±0.12 | 1.39 (0.89) | ±0.39 | 1.32 (0.82) | ±0.40 | 1.43 (0.84) | ±0.36 | 0.54 | 0.39 | 4.8E-03 |
| PANAS-13 | 2.71 (1.53) | ±0.79 | 2.65 (1.34) | ±0.58 | 1.89 (1.24) | ±0.60 | 2.70 (1.26) | ±0.54 | 0.16 | 2.06 | 0.02 |
| PANAS-15 | 3.76 (1.15) | ±0.59 | 3.57 (0.95) | ±0.41 | 3.79 (1.03) | ±0.50 | 3.30 (1.06) | ±0.46 | 0.54 | 0.38 | 4.7E-03 |
| PANAS-17 | 2.71 (1.26) | ±0.65 | 2.22 (0.95) | ±0.41 | 2.26 (1.33) | ±0.64 | 2.26 (1.01) | ±0.44 | 0.34 | 0.93 | 0.01 |
| PANAS-18 | 4.47 (0.62) | ±0.32 | 4.17 (0.78) | ±0.34 | 4.05 (0.97) | ±0.47 | 4.43 (0.79) | ±0.34 | 0.06 | 3.61 | 0.04 |
| **Neg. Emotions** | 2.85 (0.50) | ±0.26 | 2.76 (0.66) | ±0.28 | 2.67 (0.44) | ±0.21 | 2.75 (0.52) | ±0.23 | 0.49 | 0.49 | 0.01 |
| PANAS-02 | 2.94 (1.34) | ±0.69 | 2.3 (1.22) | ±0.53 | 3.21 (1.18) | ±0.57 | 3.35 (1.03) | ±0.44 | 0.15 | 2.14 | 0.02 |
| PANAS-05 | 1.53 (1.07) | ±0.55 | 1.35 (0.71) | ±0.31 | 1.95 (1.39) | ±0.67 | 1.57 (0.99) | ±0.43 | 0.67 | 0.18 | 2.3E-03 |
| PANAS-07 | 1.65 (0.86) | ±0.44 | 1.57 (0.84) | ±0.36 | 1.89 (1.15) | ±0.55 | 1.61 (0.84) | ±0.36 | 0.62 | 0.25 | 3.1E-03 |
| PANAS-08 | 3.47 (1.12) | ±0.58 | 2.48 (1.20) | ±0.52 | 2.89 (1.41) | ±0.68 | 2.91 (1.12) | ±0.49 | 0.07 | 3.47 | 0.04 |
| PANAS-09 | 1.12 (0.49) | ±0.25 | 1.17 (0.49) | ±0.21 | 1.05 (0.23) | ±0.11 | 1.13 (0.46) | ±0.20 | 0.91 | 0.01 | 1.6E-04 |
| PANAS-12 | 1.06 (0.24) | ±0.12 | 1.26 (0.69) | ±0.30 | 1.47 (0.96) | ±0.46 | 1.22 (0.52) | ±0.22 | 0.12 | 2.43 | 0.03 |
| PANAS-14 | 2.59 (1.28) | ±0.66 | 2.00 (1.45) | ±0.63 | 2.37 (1.46) | ±0.70 | 2.30 (1.36) | ±0.59 | 0.40 | 0.71 | 0.01 |
| PANAS-16 | 1.94 (1.14) | ±0.59 | 1.87 (1.06) | ±0.46 | 1.84 (1.12) | ±0.54 | 2.09 (1.20) | ±0.52 | 0.53 | 0.39 | 0.01 |
| PANAS-19 | 1.88 (1.36) | ±0.70 | 1.87 (0.69) | ±0.30 | 1.84 (1.07) | ±0.51 | 1.83 (1.03) | ±0.45 | 0.99 | 4.9E-05 | 6.2E-07 |
| PANAS-20 | 1.71 (1.21) | ±0.62 | 1.17 (0.39) | ±0.17 | 1.63 (1.01) | ±0.49 | 1.70 (0.88) | ±0.38 | 0.14 | 2.25 | 0.03 |
| **Accessibility** | 3.96 (0.54) | ±0.28 | 3.82 (0.65) | ±0.28 | 3.88 (0.70) | ±0.34 | 4.23 (0.41) | ±0.18 | 0.06 | 3.51 | 0.04 |
| WCAG-1 | 3.53 (1.07) | ±0.55 | 3.39 (1.03) | ±0.45 | 3.84 (0.83) | ±0.40 | 3.87 (0.81) | ±0.35 | 0.69 | 0.16 | 1.9E-03 |
| WCAG-2 | 4.41 (0.51) | ±0.26 | 4.17 (0.78) | ±0.34 | 3.89 (0.94) | ±0.45 | 4.48 (0.59) | ±0.26 | 0.01 | 6.48 | 0.08 |
| WCAG-3 | 4.35 (1.00) | ±0.51 | 4.04 (0.93) | ±0.40 | 4.05 (0.85) | ±0.41 | 4.78 (0.42) | ±0.18 | 0.01 | 8.22 | 0.09 |
| WCAG-4 | 3.53 (1.18) | ±0.61 | 3.65 (1.03) | ±0.44 | 3.74 (1.19) | ±0.58 | 3.78 (1.04) | ±0.45 | 0.88 | 0.02 | 3.1E-04 |

### B.3.4 Distribution of Modality Preference

**Table B.9:** Distribution of preferred modality among in-person participants and online participants.

| | **Audio** | **Audio-visual 2D** | **Undecided** | **Sum** |
|---|---|---|---|---|
| In-Person | 4 | 32 | 4 | 40 |
| Online | 4 | 36 | 2 | 42 |
| Sum | 8 | 68 | 6 | 82 |

## B.4 Participant Tables

**Table B.10:** Breakdown of participants by participation method, display modality, and digital witness.

(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 |

# C Supplementary Material to Study 2

## C.1 Questionnaire Items

### C.1.1 Questionnaire A

**Table C.1:** Translated wording of our questionnaire items to measure User Experience, Presence, Emotions, and Discomfort in the audio-visual 2D vs. audio-visual 3D study. We provide the phrasing of the corresponding items in the original questionnaires as a comparison.

| ID | Item | Equivalent in original questionnaire |
|---|---|---|
| UEQ-S-2 | I found using the interactive digital testimony easy. | The application was easy to use. |
| UEQ-S-3 | The digital contemporary witness was able to answer my questions. | The application was efficient. |
| UEQ-S-5 | I was excited by the interaction with the digital contemporary witness. | The application was exciting. |
| UEQ-S-6 | I found the conversation interesting. | The application was interesting. |
| IEQ-23 | I felt emotionally connected to the digital contemporary witness. | To what extent did you feel emotionally attached to the game? |
| IEQ-24 | I listened with interest to each response. | To what extent were you interested in seeing how the game's events would progress? |
| IEQ-27 | I sometimes forgot to use the input device. | At any point did you find yourself become so involved that you wanted to speak to the game directly? |
| IEQ-28 | I enjoyed the visual presentation of the digital testimony. | To what extent did you enjoy the graphics and the imagery? |
| IEQ-22 | I found the handling of the digital testimony easy to understand. | How well do you think you performed in the game? |
| IEQ-15 | I hardly noticed that I had to use a device to ask a question. | At any point did you find yourself become so involved that you were unaware you were even using controls, e.g. it was effortless? |
| IEQ-11 | It was as if I sat across a real person. | To what extent did you feel that you were interacting with the game environment? |
| IEQ-9 | I was aware of my surroundings. | To what extent did you notice events taking place around you? |
| IEQ-12 | I had the feeling of being in the same room with the contemporary witness. | To what extent did you feel as though you were separated from your real-world environment? |
| IEQ-2 | I was focused on the contemporary witness the entire time. | To what extent did you feel you were focused on the game? |
| IEQ-10 | I did not feel the urge to pause for a moment to take in my surroundings. | Did you feel the urge at any point to stop playing and see what was happening around you? |
| IEQ-14 | I had the feeling that the conversation with the digital contemporary witness was real. | To what extent was your sense of being in the game environment stronger than your sense of being in the real world? |
| IEQ-5 | The time has passed very quickly. | To what extent did you lose track of time, e.g. did the game absorb your attention so that you were not bored? |
| IEQ-32 | How immersed did you feel? | How immersed did you feel? |

| | | |
|---|---|---|
| CSHC3D | I was embarrassed when I had troubles phrasing my questions. | I'm embarrassed that I can't express myself well. |
| CAXM2D | I was anxious to ask the contemporary witness certain questions. | I get scared that I might say something wrong, so I'd rather not say anything. |
| CAXP2D | I was awed by the interactive digital testimony. | I get tense in class. |
| CJOA2D | I enjoyed the interaction with the digital contemporary witness. | I enjoy being in class. |
| CJOC1B | After the introductory testimony, I looked forward to asking my own questions. | I am looking forward to learning a lot in this class. |
| CAGC1D | If my question was not answered correctly, I felt frustrated. | Thinking about the poor quality of the course makes me angry. |
| SSQ-1 | How comfortable did you feel during the conversation? | I am affected by general discomfort. |
| SSQ-2 | How fatiguing was the conversation with the interactive digital testimony? | I am affected by fatigue. |
| SSQ-4 | How straining was the interactive digital testimony for your eyes? | I am affected by eye strain. |
| SSQ-8 | Did you experience dizziness or nausea during the interview? | I am affected by nausea. |

### C.1.2 Questionnaire B

Questionnaire B consisted of a single item inquiring about the participants' preferred output modality.

**Table C.2:** Binary choice between 2D and 3D.

| **ID** | **Item** |
|---|---|
| Preference | Which presentation mode of the interactive digital testimony did you prefer? |

## C.2 Interview Items

**Table C.3:** Questions asked during the fully structured interview.

| **ID** | **Item** |
|---|---|
| 1 | Which presentation method do you prefer and why? |
| 2 | What do you think of the 2D testimony? To what extent did it come close to a real conversation with a Holocaust survivor? |
| 3 | What do you think of the 3D testimony? To what extent did it come close to a real conversation with a Holocaust survivor? |
| 4 | In which display mode did you feel more comfortable and why? |
| 5 | Which display mode did you find more realistic and why? |
| 6 | Did you notice anything else during the interactive testimony that you did not mention in the questionnaire or the interview? |

## C.3 Results of Statistical Analysis

For measurements with non-normal distributions, we used paired Wilcoxon signed-rank test and its corresponding effect size r. Only the data sets for the items Presence and Emotion are normally distributed. Items IEQ-9 and CSHC3D are reverse keyed.

**Table C.4:** Full data for comparison of modality. Mean, standard deviation, 95% confidence interval. p significance, statistic (z for Wilcoxon signed rank test, t(df=50) for paired t-test), r is effect size. 2D is audio-visual 2D, 3D is audio-visual stereoscopic 3D.

| Variable | 2D | | 3D | | p | Statistic | Effect |
|---|---|---|---|---|---|---|---|
| | M (SD) | $CI_{95\%}$ | M (SD) | $CI_{95\%}$ | | | |
| **User Exp.** | 3.87 (0.57) | ±0.16 | 3.92 (0.64) | ±0.18 | 0.57 | z = 0.57 | r = 0.08 |
| UEQ-S-2 | 3.20 (1.13) | ±0.32 | 3.33 (1.19) | ±0.34 | 0.53 | z = 0.64 | r = 0.09 |
| UEQ-S-3 | 3.12 (0.93) | ±0.26 | 3.33 (0.95) | ±0.27 | 0.31 | z = 1.03 | r = 0.14 |
| UEQ-S-5 | 4.55 (0.58) | ±0.16 | 4.55 (0.64) | ±0.18 | 1.00 | z = 0.02 | r = 2.6E-03 |
| UEQ-S-6 | 4.63 (0.53) | ±0.15 | 4.47 (0.70) | ±0.20 | 0.19 | z = -1.33 | r = -0.19 |
| **Presence** | 3.26 (0.37) | ±0.10 | 3.27 (0.37) | ±0.10 | 0.15 | t = -1.48 | d = 0.04 |
| IEQ-23 | 4.00 (0.80) | ±0.23 | 4.16 (0.86) | ±0.24 | 0.21 | z = 1.25 | r = 0.18 |
| IEQ-24 | 1.67 (0.80) | ±0.22 | 2.27 (1.34) | ±0.38 | 0.02 | z = 2.43 | r = 0.34 |
| IEQ-27 | 1.57 (0.90) | ±0.25 | 1.49 (0.64) | ±0.18 | 0.53 | z = -0.66 | r = -0.09 |
| IEQ-28 | 4.31 (0.65) | ±0.18 | 4.14 (0.87) | ±0.25 | 0.35 | z = -0.94 | r = -0.13 |
| IEQ-22 | 4.10 (0.81) | ±0.23 | 4.06 (0.73) | ±0.21 | 0.83 | z = -0.23 | r = -0.03 |
| IEQ-15 | 4.31 (0.65) | ±0.18 | 4.14 (0.87) | ±0.25 | 0.35 | z = -0.94 | r = -0.13 |
| IEQ-11 | 2.02 (0.86) | ±0.24 | 2.25 (0.98) | ±0.27 | 0.09 | z = 1.74 | r = 0.24 |
| IEQ-9 | 2.86 (1.13) | ±0.32 | 2.86 (1.11) | ±0.31 | 1.00 | z = -0.01 | r = -1.7E-03 |
| IEQ-12 | 3.69 (1.12) | ±0.32 | 3.61 (1.17) | ±0.33 | 0.71 | z = -0.38 | r = -0.05 |
| IEQ-2 | 4.55 (0.58) | ±0.16 | 4.24 (1.03) | ±0.29 | 0.08 | z = -1.78 | r = -0.25 |
| IEQ-10 | 3.59 (1.17) | ±0.33 | 3.86 (1.02) | ±0.29 | 0.35 | z = 0.94 | r = 0.13 |
| IEQ-14 | 3.61 (1.02) | ±0.29 | 3.59 (1.13) | ±0.32 | 0.98 | z = 0.04 | r = 4.9E-03 |
| IEQ-5 | 2.00 (0.80) | ±0.23 | 2.00 (0.82) | ±0.23 | 1.00 | z = 0.00 | r = 0.00 |
| IEQ-32 | 3.38 (0.79) | ±0.22 | 3.56 (0.93) | ±0.26 | 0.20 | z = 1.29 | r = 0.18 |
| **Emotions** | 3.27 (0.44) | ±0.12 | 3.31 (0.54) | ±0.15 | 0.56 | t = -0.59 | d = 0.09 |
| CSHC3D | 3.41 (1.12) | ±0.31 | 3.43 (1.10) | ±0.31 | 0.79 | z = 0.28 | r = 0.04 |
| CAXM2D | 3.49 (1.07) | ±0.30 | 3.20 (1.06) | ±0.30 | 0.04 | z = -2.06 | r = -0.29 |
| CAXP2D | 1.67 (0.79) | ±0.22 | 2.27 (1.34) | ±0.38 | 0.02 | z = 2.43 | r = 0.34 |
| CJOA2D | 3.98 (0.91) | ±0.25 | 3.94 (0.83) | ±0.23 | 0.85 | z = -0.21 | r = -0.03 |
| CJOC1B | 3.86 (1.11) | ±0.31 | 3.69 (1.14) | ±0.32 | 0.44 | z = -0.78 | r = -0.11 |
| CAGC1D | 3.20 (1.11) | ±0.31 | 3.35 (1.23) | ±0.34 | 0.23 | z = 1.21 | r = 0.17 |
| **Discomfort** | 2.34 (0.41) | ±0.11 | 2.79 (0.56) | ±0.16 | 1.30E-06 | z = 4.85 | r = 0.69 |
| SSQ-1 | 4.20 (0.78) | ±0.22 | 4.08 (0.88) | ±0.25 | 0.25 | z = -1.17 | r = -0.17 |
| SSQ-2 | 1.24 (0.59) | ±0.17 | 1.60 (1.01) | ±0.29 | 0.01 | z = 2.58 | r = 0.36 |
| SSQ-4 | 1.98 (1.02) | ±0.29 | 2.32 (1.27) | ±0.36 | 0.05 | z = 1.94 | r = 0.27 |
| SSQ-8 | 1.96 (1.07) | ±0.30 | 3.14 (1.28) | ±0.36 | 1.82E-06 | z = 4.78 | r = 0.68 |

## C.4 Participant Table

**Table C.5:** Breakdown of 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 |

# List of Figures

# List of Tables