Daniel Kolb

Evaluation of the Interaction with a Digital 3D Testimony – Between Emotion and Technology

Abstract English
Conversations with Holocaust survivors are an integral part of education in German schools and universities as well as part of the German memory culture. The goal of interactive stereoscopic digital Holocaust testimonies is to preserve the effects of meeting and interacting with these contemporary witnesses as faithfully as possible. These virtual humans are non-synthetic, i.e., there exists no underlying system, which extrapolates from recorded data to synthesize and generate new answers. This means that immersion-breaking difficulties common to synthetic virtual humans, such as the audio-visual uncanny valley, can be prevented. Issues resulting from technical constraints, technological barriers to entry or errors, machine and human alike, during the design and creation of the application cannot, however, be ruled out. Therefore, I conducted a preliminary study to evaluate how people perceive this first German-speaking digital interactive 3D Holocaust testimony. I investigated how the study participants perceived the technical and semantic quality of recording and display, the difficulties in using and interacting, the accuracy and relevance of the answers given as well as the authenticity and emotiveness of the virtual contemporary witness. In this paper, I detail how the study was set up, the results of the survey, and my analysis of the data.

Abstract Deutsch
Introduction

On 8 May 2020 the 75th anniversary of Victory in Europe Day was celebrated, which had marked the end of the Second World War in Europe. The event highlighted two important issues regarding Holocaust education: on the one hand, speakers, such as Germany’s President Frank-Walter Steinmeier, warned against ‘a new brand of nationalism’, ‘hatred and hate speech’ and ‘xenophobia’ (Steinmeier 2020). President Steinmeier echoed a famous speech given by his predecessor, Richard von Weizsäcker, to mark the same occasion exactly 35 years before (Weizsäcker 1985). Both called for unity, admission of responsibility, and continued remembrance. The need for a lasting memory culture in Germany has also become more apparent in the face of ongoing antisemitism (Bundesverband der Recherche- und Informationsstellen Antisemitismus e.V. 2020), the surge of right-wing populism (Greven 2016, Lees 2018), and the trivialization of the Holocaust by elected officials (Reuters 2017, Independent 2017, Deutsche Welle 2018).

On the other hand, 75 years represent a significant length of a lifetime. Infants born in concentration camps in the last months of the war are senior citizens now and those who were teenagers and young adults at the time are currently approaching and even surpassing 100 years of age. Their advanced age increases the physical and mental toll of visiting schools, museums or memorials to recount their experiences. These face-to-face meetings and conversations with Holocaust survivors are an integral part of civic education in Germany. However, in the not too distant future, there will be no first-hand witnesses of the Holocaust left to tell their stories. The aforementioned 75th anniversary of Victory in Europe Day, as well as the anniversaries of the liberation of concentration camps like Sachsenhausen (Gedenkstätte und Museum Sachsenhausen 2020) or Dachau (KZ-Gedenkstätte Dachau 2020), gave a glimpse into this future without Holocaust survivors. Health concerns due to the COVID-19 pandemic mean that gatherings involving Holocaust survivors, who are highly at risk due to their age, are currently not justifiable.

Future generations of school students will not have the benefit of learning about the Holocaust directly from those who survived it and will have to rely on written accounts or recordings (‘Teaching the Holocaust in Germany as a Resurgent Far Right Questions It’ 2019). There is also a danger that a complete lack of Holocaust survivors and first-hand witnesses may embolden Holocaust deniers. Therefore, each passing day increases the necessity of preserving conversations with the surviving eyewitnesses as faithfully as possible. The USC Shoah Foundation undertook the first steps towards an innovative
solution as part of their Dimensions in Testimony initiative.\textsuperscript{1} They succeeded in creating the first interactive Holocaust testimonies, which are currently presented in selected museums. The Learning with digital testimonies project (LediZ),\textsuperscript{2} however, aims to incorporate interactive testimonies into German school education. This requires a stricter pedagogical focus, German-language testimonies and, eventually, a higher degree of versatility. LediZ is a cooperation between the Ludwig-Maximilians-Universität München (LMU), the Leibniz Supercomputing Centre (LRZ), and the Forever Project.\textsuperscript{3} The Forever Project previously assisted the National Holocaust Centre and Museum with their own versions of digital interactive Holocaust testimonies.\textsuperscript{4}

In this paper, I will present and discuss my preliminary study on the current state of the LediZ digital testimony project. The study focused on the capture and display quality, the interacti design, the accuracy of matching the questions and responses as well as the emotiveness and immersiveness of the chosen design.

Related Work
In 2020, Zick et al. (2020) documented a sustained high level of interest among Germany’s youth in critically examining the nation’s role in the Second World War. But their findings also show that younger Germans have less and less contact with or access to contemporary witnesses. Across all ages surveyed, 64.4 per cent had never spoken to contemporary witnesses about their experiences during the war. The authors found that those with an affinity for history education and remembrance were much less likely to have nationalist tendencies or to trivialise history. Therefore, I consider the preservation of conversations with Holocaust survivors to be a constructive part of civic education.

The first implementations of artificial dialogs with virtual humans date back as far as 1998: Marinelli and Stevens (1998) recorded an actor who was dressed as a historical figure while answering questions about that person’s life. They then categorized and indexed the video files, creating a database containing a searchable library of answer clips. Their set-up also included a speech recognition system as well as a natural language processing (NLP) model. This combination enabled users to verbalize questions, which were then analyzed for specific keywords and assigned a suitable answer video, thereby

\textsuperscript{1} https://sfi.usc.edu/dit.
\textsuperscript{2} https://www.en.lediz.uni-muenchen.de/index.html.
\textsuperscript{3} https://foreverproject.co.uk/.
\textsuperscript{4} https://www.holocaust.org.uk/foreverproject1.
simulating a conversation. The authors also used specific videos to fill the idle time between the end of an answer and the beginning of the next question, providing a livelier portrayal of the character than the use of a still image would offer. In order to deal with questions for which the program was unable to provide a suitable answer, Marinelli and Stevens established a pool of generic responses. Failure to find an answer could result from either speech recognition errors or the absence of a suitable video file. In response to the former, the virtual character asked the user to rephrase their question, while ‘out-of-bounds’ questions were met with an attempt to direct the user to another topic. This concept, consisting of an indexed collection of previously recorded response videos, speech recognition, language processing models, idle loops, and dedicated videos to handle unserviceable input, serves as a foundation for most modern interactive virtual testimonies, including LediZ.

Traum et al. (2015) successfully applied this idea to Holocaust education by adjusting a few details: Firstly, instead of actors imitating a historical figure, they filmed the real contemporary witness. Secondly, they introduced a systematic sourcing and ranking of potential topics and questions before videotaping the survivor. Thirdly, with almost two decades of technological progress since the initial work by Marinelli and Stevens, Traum et al. could rely on more efficient and capable hardware and software. This led to higher audio-visual quality capture methods and displays, a more robust speech recognition and matching process, as well as the ability to use the material in stereoscopic 3D displays.

For example, a study by Yang et al. (2012) found that those viewing videos in stereoscopic 3D felt more immersed in the subject matter, but there was an increased risk of viewing symptoms such as motion sickness. The authors note that these side effects can be lessened by increasing the distance from the display screen. Since conversations with Holocaust survivors rarely include rapid movements or changes in perspective, I find that the negative effects of using stereoscopic 3D are minimised in this context.

The use of the virtual conversation technique to explore testimonies of Holocaust survivors aims to curb an issue with Holocaust education pointed out by Gray (2014: 105): ‘One of these dilemmas is the dissemination of such an abundance of testimony material’. As an example, the author details the Tree of Testimony\(^5\) at the Los Angeles Museum of the Holocaust. The installation includes 105,000 hours, almost twelve years’ worth, of videotaped testimonies. The sheer amount of data makes meaningful search and

navigation methods, like the aforementioned voice input, essential. Frosh (Frosh 2018) puts forward a further point in support of interactive testimonies: well-designed Graphical User Interfaces (GUI) for navigating testimonies make them easier to use. He argues that simple interaction designs strengthen the involvement of the viewer. Removing technological hurdles facilitates interaction, making the testimony more accessible. A user’s subsequent unwillingness to interact may then be a moral decision on their part, stemming from an aversion to the topic, or indifference.

Since the pool of answer videos is finite and determined by the questions posed during the filming process, the potential knowledge gain of a recorded testimony is limited. For example, if the virtual witness is unable to provide a satisfactory answer to a question frequently asked by users, then the possibility of synthesizing new responses gains appeal. One possible approach could be the use of artificial intelligence to create new video and audio tracks that are based on existing recordings and combined with other available data related to the Holocaust survivor. An alternative would be the creation and animation of a model of the survivor using 3D computer graphics (CG) (Ma, Coward and Walker 2017). The project team opted against these methods, as artificial humans tend to be perceived either as abstract and unrealistic or realistic but uncanny (Timwell, Grimshaw et al. 2011). The existence of an audio-based uncanny valley, which describes eerie and unpleasant feelings in observers of imperfect artificial humans, reinforces these issues (Timwell, Grimshaw and Nabi 2015). The ongoing improvements of deepfakes, an automated method for replacing people in existing media with other people’s likeness (Thies et al. 2016), however, could create convincing synthesized answers. Yet, even without perceptive challenges undermining the effect of a virtual contemporary witness, artificially generated answers could risk the credibility, fidelity, and validity of the entire testimony. We therefore decided against amending the pool of answers.

**Implementation**

In order to create a digital copy of a living human being, the team opted to use a stereoscopic 3D recording technique. This involves displaying two slightly different video streams, one for each eye, with their perspectives offset by the mean interpupillary distance. It provides a more detailed and immersive visual experience due to the added dimension and depth. This approach, however, increases the technical requirements for displaying the video data, as it requires installations capable of delivering separate pictures to each eye. A common solution is the temporal or polarized overlay of both
streams on a single screen. 3D glasses equipped with an active shutter functionality or polarization filter are then used to separate the images for each observer, depending on which system is used to play the video. The doubling of the video information is a further burden, since it also results in a duplication of the data load. This can lead to bottlenecks and delays, depending on the quality of the connection to the data location and the decoding time on underperforming hardware. If these requirements prove to be too limiting in practical use, it is, however, always possible to fall back on conventional 2D display techniques. The two video streams are captured separately during the process of filming, which makes it possible to readily convert the films from 3D into 2D by filtering the streams intended for one eye. We intend to future-proof digital testimonies by using technological standards that are already well established, albeit not yet present in average households or exhibitions. Consequently, even though the processed video data we are presently employing uses a resolution of 1080 HD and 24 Hz frame rate per eye, we originally captured the raw video data with a resolution of 6K HD and 50 Hz video frequency per eye with the help of Bright White Ltd. Once displays supporting higher resolutions and bandwidth become more widely used, we can adjust and upgrade the quality of the videos with little effort.

Furthermore, we preserve all captured, generated and processed media data in a long-term tape archive at the Leibniz Supercomputing Centre (LRZ). For additional safety and redundancy, the content of the archive is mirrored at the Max Planck Computing and Data Facility (MPCDF). The LRZ also hosts the processed video files, which can then be streamed directly to the application and installation. This solution eliminates the need to provide the data locally.

---

6 [http://www.brightwhiteltd.co.uk/](http://www.brightwhiteltd.co.uk/).
7 [https://www.mpcdf.mpg.de/](https://www.mpcdf.mpg.de/).
The project team chose the Centre for Virtual Reality and Visualisation (V2C), also part of the LRZ, as the prototype installation and testing facility. With a 3D-capable 6m wide and 3.15m tall powerwall and dedicated seating for groups of up to 21 people, the centre enabled us to test, demonstrate and evaluate the intended digital testimony. This is also the environment in which I conducted my survey (see Figure 1). I provided each participant with polarized 3D glasses so that they could experience the contemporary witness video in 3D. I also supplied each test person with a dedicated smartphone as a mobile input device for questions. The application on the smartphone has two states: idle and recording. Pressing anywhere on the screen switches from idle to recording, while lifting the finger returns the input device to the idle state. A change in background color from black to blue visually confirms the state change to the user (see Figure 2). The captured voice input is then sent to a continually trained NLP system, which aims to extract the core of the question. If the system successfully finds a matching answer, the corresponding video is presented to the viewer. If no suitable answer for the input is found, the virtual contemporary witness asks the user to rephrase their question.

This study focusses on the interaction with the interactive virtual 3D testimony given by Abba Naor. His virtual testimony is presently more advanced as the project team had had more time to thoroughly train the underlying NLP system.
Results and Discussion

I surveyed 46 participants for the study (see figure 3) over a period of two weeks at the end of February and beginning of March 2020. The imminent threat of SARS-CoV-2 and the resulting federally imposed restrictions on public life meant that further surveys scheduled for the spring and summer of 2020 had to be abandoned. Of the 46 people surveyed, 43 per cent identified as female, 57 per cent identified as male, and none as non-binary. Exactly half of the participants were under 26 years old, representing the average age range of school or university students (Statistisches Bundesamt 2019). More than 90 per cent reported at least a fair amount of interest in topics related to the Holocaust and, thus, have previously investigated these topics during their spare time and of their own volition (see figure 4). Furthermore, figure 5 shows that two out of every seven people surveyed had personally interacted with one or more Holocaust survivors, either within their own family, among their acquaintances, or as part of their education. Before engaging with the digital Holocaust survivor, only one in seven had already experienced digital historical testimonies, for example, in the form of Instagram stories.

I provided each participant with a short printed summary of the most significant stages in Abba Naor’s life. I chose a short and compact handout over a more detailed insight into his life story that was available in the form of a 41-minute video narrated by Abba Naor himself. Using the handout rather than the introductory video gave the participants more time to ask questions as well as providing an easily accessible, structured overview of possible topics of conversation. Several participants, however, later noted that they would have preferred the more detailed account by the digital witness. They remarked...
that an introductory narration by the virtual conversational partner would have put them more at ease and would also have supplied them with a wider range of ideas for potential questions.

After a brief explanation on how to use the system, all participants spent 25 to 35 minutes interacting with the digital testimony. I did not interact with, interfere with or limit the test persons during this process, except when they required technical support or asked for adjustments, e.g. lighting or volume, to be made. At the end of their interaction period, I asked each participant to report their impressions and reception of the digital interactive testimony by completing a questionnaire. For each resulting set of data, I also provide the cardinality n of the respective set, as not all participants answered every question in the questionnaire.
The first part of my evaluation addressed the design of the app on the mobile input device. The results, illustrated in figure 6, show a high degree of satisfaction among the users. This refers particularly to the simplicity of the interaction process and the GUI. Three participants expressed dissatisfaction with the stability of the application. In these cases there was a delay between the end of the user’s input and the start of the reply video by the digital testimony. The application itself, however, was unlikely to be the cause of these issues, it was most likely due to another component in the system such as the streaming server, the matching system or the network connecting the modules. I then examined the participants’ opinion of the capture quality and the audio-visual features of the installation. The responses were predominantly positive (see figure 7). The respondents seemed particularly satisfied with the audio capture and playback quality and with the acoustics of the V2C. They also did not appear to be perturbed by the noise of other participants present. A significant reason for this is the respectful and reverent behavior of the participants, something also observed in the presence of real-life Holocaust survivors. Some respondents registered small issues with the intelligibility of the answers. I attribute this to Abba Naor retaining a few peculiarities of his native languages, Lithuanian and Hebrew, when speaking German. Both video qualities, capture and playback, were rated slightly lower than the audio qualities, but were still highly satisfactory overall.
The final section of the questionnaire dealt with the answers given by the digital testimony. A dedicated query regarding the perceived matching accuracy showed that the
responses given by the virtual contemporary witness were suitable to the majority of respective questions (see figure 8). The overall matching accuracy during my study was 63 per cent, which I consider to be insufficient. The project team aims for an average matching rate of 90 per cent, which would guarantee a more satisfying conversation. During previous randomised checks and self-tests, the match rate of the testimony reached and exceeded this threshold. However, this study is the first to evaluate the virtual testimony in an entirely unconstrained setting with test persons devoid of any prior experience. In response to this finding I carefully inspected the source material and the matching system and. With the help of detailed feedback from the participants, was able to identify previously undetected faults and inconsistencies in the matching system, which were the result of human error during the building phase. Additionally, I found that only 15 per cent of the participants were satisfied with the content of the answers given, while 22 per cent were dissatisfied (see figure 9). However, 30 per cent reported that the answers given matched the essence of their questions, while 15 per cent felt the answers did not fit the questions. More than 80 per cent replied that the answers included further details not directly relevant to the question but 88 per cent considered these details to be of great interest. In terms of the length of the answers given, 80 per cent of the participants were against shorter answers, yet there was ambivalent feedback to a proposal for longer answers. Thus, I can conclude that the captured footage of Abba Naor is suitable for this format, however, this may, in part, be due to the fact that Abba Naor has spent many years answering questions about his life and refining his answers. Seven-in-ten people surveyed confirmed that they perceived the emotions of the Holocaust survivor, which were evoked by his recollection. Moreover, 41 per cent said they felt like they were talking to and interacting with a real human being. When asked about the content of the answers given by the digital testimony, 26 per cent reported having learned a substantial amount of new information on the Holocaust, while 28 per cent reported little to no new insights.
Figure 8 – Elaboration on the accuracy of the automated matching of answers to spoken questions (n=45)

![Matching accuracy chart]

Figure 9 – Satisfaction of participants with the content and accuracy of the answers as well as the emotiveness of the interactive component (n=46 with the exceptions of n[interesting additional details]=45, n[knowledge gain]=45)

![Quality of the dialogue chart]
I asked the study participants to rank the significance of four main characteristics of the interactive stereoscopic digital Holocaust testimony: amount of available answers, 3D effect, visual fidelity and ease of use. The results are detailed in figure 10. The aspect classified as most important by far, was a large pool of possible responses. The feature considered least important of the four was the 3D effect. Nevertheless, it also received the second highest amount of ‘most important’ votes, along with ease of use. The test persons viewed both visuals and usability with ambivalence. These results raise the question of the practical benefit of the stereoscopic 3D effect. However, the survey did not include a comparative study on the measurable difference in effect on the viewer between a 2D interactive testimony and a 3D interactive testimony. Therefore, it is not possible to answer this question in a scientifically sound way.

![Ranking importance of features](image)

Figure 10 – Subjective prioritization of four features: Quality of responses, 3D effect, visual quality, and handling (n=42)

When prompted to assess the interactive testimony as a whole, 60 per cent of the participants gave it the school grade ‘2’, which translates to the letter grade ‘B’ (see figure 11). The average grade amounts to 2.47 or ‘B-’. While these responses highlight the need for improvement and refinement in some components of the system, it also describes an overall satisfactory and meaningful experience. This is further supported by 88 per cent of the participants being in favor of interactive stereoscopic digital Holocaust testimonies being used in museums or schools on a regular basis. I provide a breakdown of this feedback in figure 12.
Conclusion and Future Work

This paper details the first assessments of the first German-speaking digital interactive Holocaust testimony within the scope of a preliminary user study. I focused on evaluating the chosen technical implementation: the quality of the user interaction, the audio-visual fidelity of the recordings and their playback, as well as the accuracy of the system in matching answers to spoken questions. My results can assist in the design, implementation, and evaluation of further digital testimonies as well as the refinement of existing virtual contemporary witnesses.

I found that the user interface and the interaction itself were regarded as easy to use and understand. This characteristic is crucial for an application aiming for acceptance and adoption by people of different backgrounds, ages, or levels of education. I also found
that the chosen implementation, including the quality of the media data, was positively received. Overall, the study participants found the answers interesting. However, I have identified shortcomings in the current state of the matching system. Improvements and adjustments in this sector could lead to higher accuracy and, therefore, a more satisfying experience. Nevertheless, the present version of the digital interactive Holocaust testimony of Abba Naor was well received, and there was strong support for its use in educational and cultural facilities.

During the evaluation I identified the need for a separate study focusing on the advantages and disadvantages of stereoscopic 3D testimonies as opposed to monoscopic 2D testimonies in the context of learning, immersion and emotiveness. Similarly, a study comparing the effect of digital testimonies and interaction and conversation with real Holocaust survivors would help to quantify their differences and respective benefits. Since this study only focused on the digital testimony of Abba Naor, I also intend to conduct a similar investigation of the second digital testimony, which focuses on the life of Eva Umlauf. Due to the shared history and similar issues regarding Holocaust education in Austria (Mittnik, Lauß and Hoffmann-Reiter 2020), my findings could moreover be applied and subsequently evaluated in Austria and possibly also other German-speaking countries.
References


Yang, Shun-nan et al. ‘Stereoscopic viewing and reported perceived immersion and symptoms’, in: *Optometry and vision science* 89, 7 (2012), 1068-1080.