A Cloud-based User-Centered Time-Offset Interaction Application

Alberto Chierici, Tyeece Hensley, Wahib Kamran, Kertu Koss,
Armaan Agrawal, Erin Collins, Goffredo Puccetti and Nizar Habash
Computational Approaches to Modeling Language (CAMeL) Lab
New York University Abu Dhabi, UAE
{alberto.chierici,nizar.habash}@nyu.edu

Abstract

Time-offset interaction applications (TOIA) allow simulating conversations with people who have previously recorded relevant video utterances, which are played in response to their interacting user. TOIAs have great potential for preserving cross-generational and cross-cultural histories, online teaching, simulated interviews, etc. Current TOIAs exist in niche contexts involving high production costs. Democratizing TOIA presents different challenges when creating appropriate pre-recordings, designing different user stories, and creating simple online interfaces for experimentation. We open-source TOIA 2.0, a user-centered time-offset interaction application, and make it available for everyone who wants to interact with people's pre-recordings, or create their pre-recordings.

1 Introduction

Stories are the most pervasive medium people use to make sense of themselves and the surrounding world. In the past twenty years, social media development has enabled people to document narratives of their daily experiences online in unprecedented measure (Page, 2013). TOIAs explore the next evolution of narrative sharing devices exploiting advances in artificial intelligence, as well as peoples' increasing comfort with virtual communication methods, to simulate face-to-face interactions between humans and previously-recorded videos of other humans.

Democratizing time-offset interaction has many challenges, including but not limited to: creating appropriate videos, robust human-computer interaction (HCI) design, and lack of intuitive online interfaces. We open-source TOIA 2.0, a user-centered TOIA, and make it available for everyone who wants to interact with people's pre-recordings, create their own, and researchers in dialogue, machine learning, and HCI to gather original datasets.

2 Related Work

The usefulness of TOIAs has been demonstrated in many practical scenarios. For example, for keeping historical memories (Traum et al., 2015), job interview practice for young adults with developmental disabilities,[1] and building digital humans across different industries. Storyfile, Typeform's videoask are some examples of commercial applications.[2] TOIAs may also be reminiscent of virtual assistants like Siri and Alexa and digitally animated characters like Digital Humans,[3] and Soul Machines;[4] however, these are not authentic representations of human beings which is TOIAs' goal.

The general public cannot afford current TOIA deployments due to their high production costs: creating a character (aka avatar) may require prerecording about 2,000 video answers (Nishiyama et al., 2016; Jones, 2005). Chierici et al. (2020) proposed a more streamlined avatar development process, but it is still impractical for the everyday user: it involves transcribing and recording conversations based on brainstormed plausible utterances. Their work resulted in creating more than 400 prerecordings and manual annotations that took several days. Research into time-offset interactions needs to generalize and to streamline the avatar development process to make a mass use system. A first attempt made by Abu Ali et al. (2018) goes towards this direction and includes the possibility to chat with the avatars in different languages. Their system implementation is not simple to use because it has two separate, non-communicating components for recording videos and interacting with them. It also requires local installation and does not support multiple users.

[1] https://ict.usc.edu/prototypes/vita/
[2] www.storyfile.com, www.videoask.com
[3] www.digitalhumans.com
[4] www.soulmachines.com

Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue, pages 265–268

3 Design Principles

We designed TOIA 2.0 so that the time-offset interaction feels like a natural interaction to all users: the user who interacts with other people's pre-recordings (henceforth, *interactor*), and the user who records their own narratives (henceforth *TOIA maker*). As per best practices in UX design, each step of interaction with the interface should minimize the user's cognitive load and be psychologically satisfying. We followed the human-computer interface heuristics established by Nielsen (1994). Some design decisions to reach this goal included: the creation of a common visual framework across the whole system for all types of users; use of familiar layouts and vocabulary; creation of affordances by auto-generating suggested questions to answer and to suggest particular kinds of interactions; providing psychological satisfaction with profiles that emphasize social interaction; giving users agency and autonomy over aspects of their TOIA experience; and providing reassurance with repeated language and layouts.

An important extension to the previous work by Abu Ali et al. (2018) is the introduction of a social network aspect. This does not just serve the goal of creating a community of TOIA makers, but also provides helpful feedback to them including what additional questions are asked, what answers are liked or disliked. Another extension is giving the TOIA maker control over which recordings are playable individually, and as part of *streams* (i.e. collections or albums) that define different contexts and intentions for the interactions.

4 TOIA 2.0 System Architecture

The cloud-based centralized TOIA 2.0 system consists of six components, broadly speaking, that are interconnected via a web server and component APIs (Figure 1). In the rest of this section, we discuss the back-end components of TOIA 2.0. We then proceed to discussing the user interface which is the highlight of this demo paper in Section 5.

TOIA Database The central repository of the TOIA 2.0 system is a relational database management system that stores all user information from the TOIA makers' name, password, biographical description, and language preferences, to links to, and meta-data of, all their video recordings, and their stream organization, as well as suggested questions.

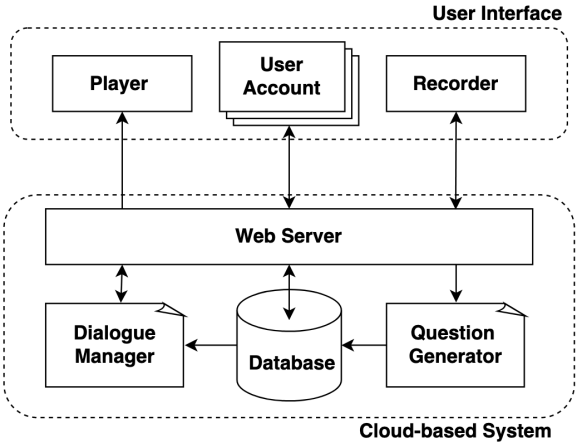

Figure 1: TOIA 2.0 system architecture

Dialogue Manager The Dialogue Manager is implemented as microservice that takes as input the ASR output of last utterance of the interactor conversing with a specific TOIA stream. It returns the video ID of the video recording containing the response. We use Chierici and Habash (2021)'s *BERT q-q* retrieval index as implemented by Haystack.[5]

Question Suggestion One of the most crucial steps in creating a TOIA is building a knowledge base consisting of questions and recorded video answers (pre-recording). Although users have complete creative freedom when it comes to creating pre-recordings, it might still be tedious and challenging to come up with and record hundreds of responses while trying to predict the hypothetical paths of future conversations.

We use transformer models such as GPT-2 to generate personalized suggestions (Radford et al., 2018; Mishra et al., 2020). The generative pre-trained transformer fine-tuned on dialog-specific data outputs a collection of question suggestions, given the pre-recordings maker's profile settings and the history of already recorded question-answer pairs. In that way, new personalized questions can be dynamically generated and suggested at any given point during the avatar creation process to any user given the same base model.

Web Server The web server is the main component that connects all the TOIA 2.0 parts. It manages information flow from front-end to back-end and vice versa.

[5] haystack.deepset.ai/

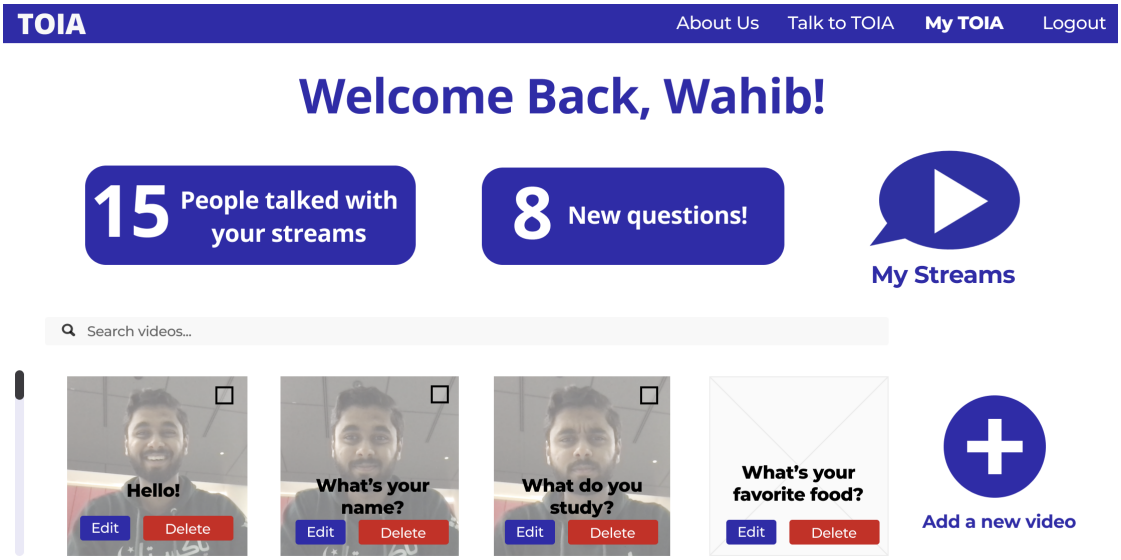

Figure 2: TOIA maker User Account

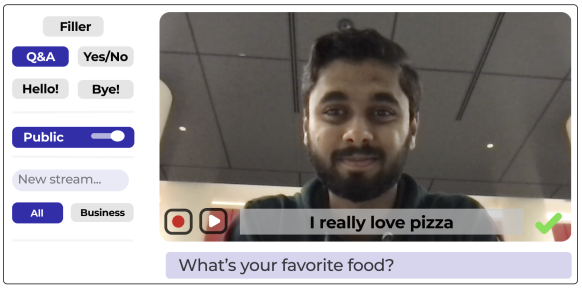

Figure 3: A TOIA maker recording a video answer to the question "what's your favorite food?"

5 User Interface

The user interface (UI) comprises a home page, login and sign-up screens, the recorder page, the player, and various menu and navigation aid items. While the TOIA 2.0 main page allows browsing profiles of different public TOIA streams, all users (TOIA makers and Interactors) must register and have an account to be able to make their own TOIA streams or converse with public TOIA streams. A full walk through the user flow is available on the application page.[6] We next present specific components.

User Account Figure 2 presents the user account for a registered TOIA maker. The user is greeted by a message and shown statistics about the number of people that interacted with their TOIA along with a notification of any new questions that were automatically generated by the system or identified though failed interactions (e.g., when an answer was not found with high confidence).

Streams To the mid right of the figure, there is a link to a management page for the streams associated by the TOIA maker. Each stream can get its own profile that specifies the functional purpose of the stream, e.g., a *business* stream may target job interviews, while a *family stream* can focus on sharing family histories. Streams can be made public or be only shared with specific users. All viewable streams can be accessed on the *Talk to TOIA* page (see Figure 4).

Videos The bottom half of Figure 2 shows a collection of recorded videos, and some entries with questions and no recordings. New questions are presented as videos that have no content and are waiting to be filled. The TOIA maker can delete existing video entries or suggested questions, or click or *Edit* to record a response or change an existing response. The TOIA maker can also create a completely new video by clicking on the big plus sign to the bottom right of the figure. Edit and Add actions will take the TOIA maker to the recorder view. The checkboxes at the top right corner of the video allow the TOIA maker to select a number of recordings and assign them to a stream. All videos in the TOIA maker's account can be filtered for display using keyword search.

Recorder Figure 2 shows the recorder page, which is opened when a TOIA maker chooses to edit a video or create one from scratch. The main

[6] http://toia.camel-lab.com/

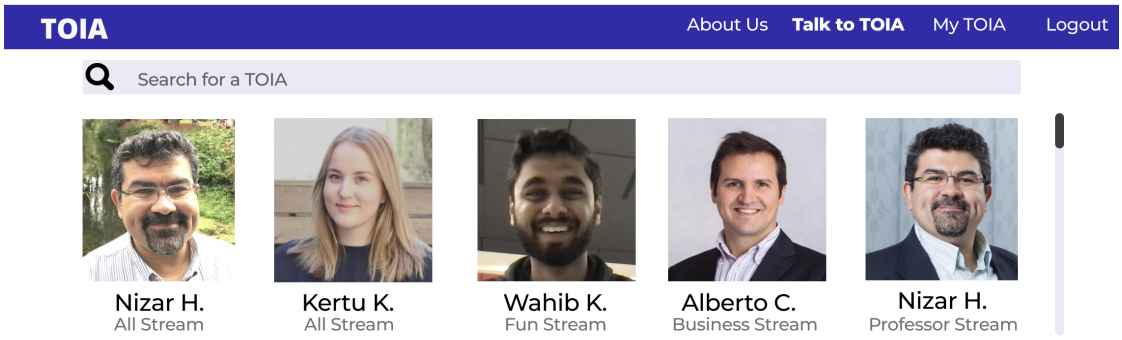

Figure 4: A view of a number of public TOIA streams that are ready for interactors to converse with.

component of the recorder page is recording screen connected to the user's camera. The question (or prompt) is at the bottom of the screen. Automatic speech recognition output is shown to the TOIA maker, and they have the option of editing the text. To the left of the recording screen are a set of labels. First are the video recording type, e.g., *filler video* or *Question & Answer video*. Next is a toggle for making the video public. Finally a list of available streams. All videos are automatically assigned to the *All* stream. A video can be assigned to more than one stream. Creating a stream is as easy as writing a new stream name. After recording a video, the TOIA maker can test play it, rerecord it, or delete it.

The Player The player interface can be accessed through the *Talk to TOIA* page, which lists all the publicly viewable streams which the interactor user has access to (see Figure 4). The player interface is intentionally as simple as a *Facetime* or *Skype* interface: the interactor speaks, and the player plays an appropriate response from the list of stream videos as determined by the dialogue manager.

6 Conclusion and Future Work

We presented TOIA 2.0, a user-centered time-offset interaction application system, which we plan to make publicly available to users – TOIA makers and interactors – as well as researchers interested in TOIA systems.

In the future, we plan to increase the robustness of the system and its membership. We are also considering additional enhancements to both the TOIA maker and interactor's experience, e.g., providing semi-automatically generated videos or guessing answers on behalf of a TOIA maker given their history. In all cases, we keep the TOIA maker in complete control of their recordings and streams.

References

Dana Abu Ali, Muaz Ahmad, Hayat Al Hassan, Paula Dozsa, Ming Hu, Jose Varias, and Nizar Habash. 2018. A bilingual interactive human avatar dialogue system. In *Proc. of SIGdial Meeting on Discourse and Dialogue*.

Alberto Chierici and Nizar Habash. 2021. A view from the crowd: Evaluation challenges for time-offset interaction applications. In *Proc. of the Workshop on Human Evaluation of NLP Systems*.

Alberto Chierici, Nizar Habash, and Margarita Bicec. 2020. The margarita dialogue corpus: A data set for time-offset interactions and unstructured dialogue systems. In *Proc. of Language Resources and Evaluation Conference*.

Karen Spärck Jones. 2005. Some Points in a Time. *Computational Linguistics*, 31(1).

Shlok Kumar Mishra, Pranav Goel, Abhishek Sharma, Abhyuday Jagannatha, David Jacobs, and Hal Daume. 2020. Towards automatic generation of questions from long answers. *arXiv:2004.05109*.

Jakob Nielsen. 1994. Enhancing the explanatory power of usability heuristics. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*.

Masashi Nishiyama, Tsubasa Miyauchi, Hiroki Yoshimura, and Yoshio Iwai. 2016. Synthesizing realistic image-based avatars by body sway analysis. In *Proc. of International Conference on Human Agent Interaction*.

Ruth E. Page. 2013. *Stories and Social Media: Identities and Interaction*. Routledge.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training. *Preprint short-url.at/fzBC4*.

David Traum, Andrew Jones, Kia Hays, Heather Maio, Oleg Alexander, Ron Artstein, Paul Debevec, Alesia Gainer, Kallirroi Georgila, Kathleen Haase, et al. 2015. New dimensions in testimony: Digitally preserving a holocaust survivor's interactive storytelling. In *Proc. of International Conference on Interactive Digital Storytelling*.