Towards Glanceable On-Demand AR Conversation Visualization

Shanna Li Ching Hollingworth*
University of Calgary

Wesley Willett†
University of Calgary

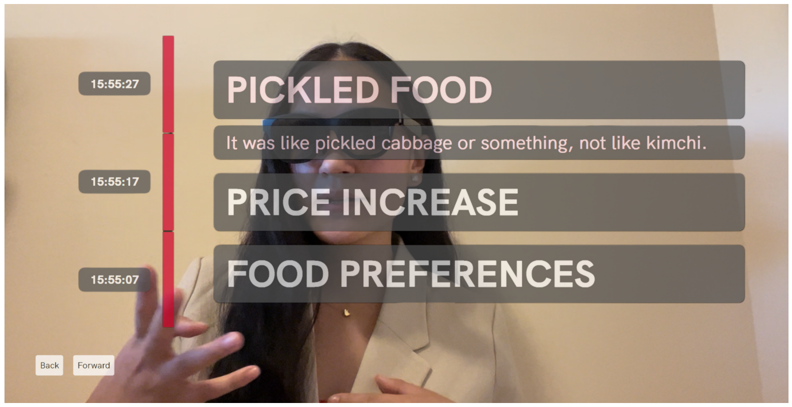

Figure 1: A screenshot of the AR timeline prototype (XReal does not currently support direct screenshotting, so the visualization is superimposed on the image)

Abstract

We explore the potential for glanceable real-time conversation timelines in the Augmented Reality (AR) space, with a focus on lightweight and non-intrusive design. In recent years, AR has become an increasingly popular option for conversational support, leading to many contributions in the field. However, there has been less focus on developing lightweight, non-intrusive, on-demand conversation visualizations in AR. It is important that visualizations don't distract from the conversation partner or the conversation itself. Furthermore, the information displayed may not always be relevant to the user at every moment. Therefore, we explore methodologies to dissect and visualize conversation timelines in real time, providing glanceable summarizations and breakdowns of conversations which are easily summoned and dismissed. In our prototyping, we use a combination of Large Language Models (LLMs) and Natural Language Processing (NLP) techniques to recognize and classify changes in topic. To display the visualization, we use the XReal sunglasses as a lightweight unobtrusive headset which could reasonably be worn in everyday life. We identify a unique set of visualization challenges and opportunities related to minimalist conversation visualizations. Following these explorations, we discuss our findings, and provide our initial reflections on various visualization techniques.

Index Terms: Augmented Reality, Conversation Visualization.

1 Introduction

The experience of getting halfway through a conversation and suddenly realizing that the topic has deviated far from the initial subject, subsequently exclaiming, "How did we get here?!", is familiar to most. The natural flow of a conversation and the transition from one topic to another is unique every single time. Where two different groups starting with the same conversation prompt end up after 5 minutes is often entirely different and greatly depends on the participating individuals, as anyone who has ever participated in an organized breakout room conversation might have observed. Understanding how topics shift in conversation can offer insights into its broader context. However, pausing to recall the conversation's path can disrupt the natural flow, potentially derailing it entirely.

With the recent advances in lightweight Augmented Reality(AR) technology, and its ability to display information while keeping the user centered in the real world, it has become increasingly popular as a tool for analyzing and displaying information about live speech in real time [1]. However, AR also has great potential for introducing more digital clutter, making it increasingly difficult for individuals to fully engage with the real world, and maintain human connection with those around them. For example, Nguyen et al. [3] found that AR topic prompts were only beneficial when timely, as they could otherwise disrupt the conversation. Rivu et al. [4] and Liu et al. [2], who worked on gaze-driven AR systems and visual captioning in AR, respectively, emphasized minimizing distractions to maintain conversational flow. Thus, our research focuses on creating a lightweight, non-intrusive AR interaction scheme to visualizes conversation content on demand. This ensures users can access relevant information when needed, while otherwise remaining fully engaged with those around them.

2 Prototyping AR Conversation Visualization

While it is clear that AR has the potential to be a powerful and versatile tool for conversation support, it is vital that these tools are designed with the goal of fostering human connection. We prototype an on-demand tool for real-time conversational timeline visualization that minimizes gaze aversion, and uses a minimalistic design.

To ensure this, we have identified five specific design goals:

1. Lightweight visualization on demand
2. Limiting unnecessary distraction and interaction
3. Consistent summarization as text is added
4. Providing access to the raw text
5. Prioritizing privacy and consent in data collection

These goals were decided upon after many detailed discussions and brainstorming sessions within the research team and the wider

*e-mail: shanna.hollingwor1@ucalgary.ca
†e-mail: wesley.willett@ucalgary.ca

research group. These discussions and goals were also informed by various findings and limitations in the related works, particularly goals 1 and 2 regarding non-distracting design (See Section 2). The prototype should visualize conversation at multiple levels. We have currently chosen to create visualizations in time blocks of 10 seconds, 30 seconds, 1 minute, and 5 minutes, as well as by topic as defined by dialogue segmentation techniques. However, we must further evaluate which levels provide the most information.

For our initial prototype, we selected XReal sunglasses for their lightweight, unobtrusive and sleek design. We used a simple HTML webpage with D3 for quick design prototyping. The OpenAI Whisper API handled speech-to-text, and we analyzed and classified transcripts using the OpenAI GPT-4o API and traditional NLP methods, including the NLTK Dialog Segmentation library.

Figure 1 shows our baseline visualization with three 10-second conversation blocks. The bar on the left represents the duration each topic was discussed, split into equal parts for 10-second segments. For topic-based modes, the bar displays the percentage of time spent on each topic. Time stamps on the left mark the ending time of each conversation block, with the top being the most recent. Our current interaction scheme lets users select a topic to view a detailed 'most representative sentence' from the transcript, as chosen by GPT-4o. Users can then 'zoom in' or 'zoom out' on this sentence to explore details from different time blocks.

3 Design Challenges

We have begun to explore a wide range of glanceable and lightweight visualization designs, as part of this work. Based on our initial experiences we highlight a variety of design challenges and opportunities for glanceable AR conversation visualizations.

3.1 Segmentation and Summarization of Live Speech

Human conversation is inherently unstructured and difficult to visualize, with topics often drifting unpredictably. Particularly, when analyzing real-time conversations, the best places to segment text for topic separation often only become evident after the fact. In our experiments, GPT-4o was highly successful in segmenting and classifying topics in a 30-minute transcript. However, it struggled with smaller, real-time transcript chunks, often failing to recognize topic changes and causing the context window to increase exponentially. While refining prompt engineering could improve results, traditional NLP methods might be more reliable. A viable approach may be implementing dialog segmentation with Python's NLTK library and then applying OpenAI for classification.

3.2 Balancing Visualization Density and Complexity with Glanceability

In prototyping, we've faced challenges in displaying extensive data with minimal visualization. We need to balance useful history duration with the number of topics that can be displayed legibly. We also consider which extraneous information, such as time spent on topics or speaker identity, adds value without unnecessary complexity. Another possible encoding involves creating different visualization modes at varying time intervals, such as every 10 seconds, 5 minutes, or over an hour, to summarize conversations at different levels of detail. We should consider including information on model certainty for transparency, but this might clutter the interface without being useful to users. Currently, the raw text context displayed in the prototype helps contextualize the LLM's classification choices.

3.3 Exploring Placement and Gaze

Placement and gaze considerations offer various visualization opportunities. Nguyen et al[4] found participants felt uncomfortable diverting their eyes to view prompts. To address this, our current visualization centers the timeline in the user's field of view to avoid obvious gaze aversion. Although this may partially obscure the partner's face, it shouldn't be a significant issue due to the brief display time. Further evaluation is needed to compare this with other approaches. Exploring these visualizations on other lightweight devices like smartwatches could be interesting and beneficial.

3.4 Exploring Minimalist Interactions for Summoning and Navigating Conversations

Interactions with the visualization should be lightweight and subtle, and gestures should trigger device responses while feeling natural in conversation. Approaches we have considered include using XReal hand tracking to register micro gestures and hand movements, though it may be too noticeable for conversation partner. Gaze-based interactions are not supported by XReal, but tracking gaze movements could be a good way of interacting with visualizations. Another consideration is the implementation of quasimodes, for extremely intentional visualization display.

3.5 Privacy

During prototype design, we identified privacy and consent as major challenges. Conversational tools might risk recording and analyzing speech without users' knowledge. One approach is to differentiate voices locally and only save and visualize speech from the tool's user. However, the usefulness of visualizing only part of the conversation is thus far unclear. We also considered integrating consent protocols requiring verbal opt-in from participants before recording their speech. Additionally, minimizing conversation history retention can reduce the risk of misuse of recorded transcripts.

Moving forward, we will further explore, dissect, and evaluate these design opportunities, with the ultimate goal of identifying a design space for glanceable, lightweight, on-demand AR for everyday conversation visualization. This prototype is envisioned to be part of a larger set of lightweight tools designed for conversation support. Other further research opportunities would involve identifying other areas in everyday conversation that would benefit from AR visualization, and applying what was learned from this prototype to these further explorations.

Acknowledgments

We would like to acknowledge Victoria Wong, who helped create mockups for the second iteration of our design and assisted in prototype development and styling. This work was supported in part by Alberta Innovates and the Canada Research Chairs Program.

References

[1] I. Hirskyj-Douglas, A. Kantosalo, A. Monroy-Hernández, J. Zimmermann, M. Nebeling, and M. Gonzalez-Franco. Social ar: Reimagining and interrogating the role of augmented reality in face to face social interactions. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '20 Companion, p. 457–465. Association for Computing Machinery, 2020. doi: 10.1145/3406865.3418585 1

[2] X. B. Liu, V. Kirilyuk, X. Yuan, A. Olwal, P. Chi, X. A. Chen, and R. Du. Visual captions: Augmenting verbal communication with on-the-fly visuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, 2023. doi: 10.1145/3544548.3581566 1

[3] T. T. Nguyen, D. T. Nguyen, S. T. Iqbal, and E. Ofek. The known stranger: Supporting conversations between strangers with personalized topic suggestions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 555–564. Association for Computing Machinery, 2015. doi: 10.1145/2702123.2702411 1

[4] R. Rivu, Y. Abdrabou, K. Pfeuffer, A. Esteves, S. Meitner, and F. Alt. Stare: Gaze-assisted face-to-face communication in augmented reality. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Adjunct. Association for Computing Machinery, 2020. doi: 10.1145/3379157.3388930 1