

Synchronized Realities: Towards Magic Mobile Experiences through Aligned AR

Jan Henry Belz
Dr. Ing. h.c. F. Porsche AG
Stuttgart, Germany
jan_henry.belz@porsche.de

Abstract

In virtual reality environments, the alignment of perceptual modalities is crucial for immersion and presence. In the AR domain, it is difficult to create such alignments because elements in the physical world are often beyond the user's control. However, recent advances in generative AI enable on-demand content creation, enabling highly reactive AR experiences. Combined with contextual information about the physical world, it has become possible to design experiences that seamlessly align with the user's environment. In this reflection paper, I emphasize the importance of "synchronized" realities for context-aware AR experiences, particularly in mobility scenarios. I present several examples of existing synchronized experiences and examine their commonalities and distinctions. Finally, I discuss opportunities and pitfalls of synchronizing AR experiences with the physical world.

ACM Reference Format:

Jan Henry Belz. 2026. Synchronized Realities: Towards Magic Mobile Experiences through Aligned AR. In *Proceedings of CHI 2026 Workshop on Next Steps for Augmented Reality On-the-Move: Challenges & Opportunities*. ACM, New York, NY, USA, 3 pages.

1 Introduction

The alignment of perceptual modalities is a core requirement for immersive virtual reality (VR) scenarios [10]. In virtual reality, the entire world can be controlled by the designer or the user; however, this is not possible in augmented reality (AR) scenarios. Physical elements of the real world can only be accessed through embedded hardware, such as in smart home environments (ubiquitous computing) or, in mobility scenarios, in-cabin technology (e.g., cars or buses) [14]. Therefore, the AR experience must adapt to the physical world to provide a seamless, integrated experience [4].

Recent advances in artificial intelligence (AI) research enabled almost human-like zero-shot interpretation of user context (or at least a mimicked version of it) [15]. Natural-language interaction with multimodal perception models enables automated description of a user's environment in "human-readable" ways, e.g., to capture the "vibe" or atmosphere of a moment. Moreover, generative AI allows the creation of customized media content that can be presented to users on demand. For example, information such as images and videos relevant to the current situation can be queried,

generated, and displayed in real time. When these methods are combined meaningfully, the user's current situation can be analyzed in depth, and an appropriate user experience can be delivered [8].

While AR generally refers to the visual overlay of information [1], mixing and augmenting reality with contextual information can occur across multiple modalities (e.g., audio augmented reality (AAR) [3]). This offers wide opportunities, particularly for mobile applications, such as adaptive music experiences [9]. Since mobile scenarios involve rapidly changing contextual information, mobile AR applications need to be even more reactive than in static environments. Therefore, to perfectly align the digital experience with the physical reality,

The resulting "synchronization" of the digital and physical realities offers enormous potential for seamless, truly ubiquitous user experiences. Through meticulously synchronized experiences, technology can feel like magic, supporting the user in their current activity with zero friction. In this reflection paper, I want to point out the potential of "synchronized reality". I present existing work that proved the effectiveness of such synchronizations, including my own contribution to this area. In three example scenarios, I illustrate commonalities and core aspects of synchronizing AR experiences to the user context. Finally, I discuss opportunities and pitfalls for designing such experiences.

2 Example Scenarios

Synchronized music. A commuter leaves the office at dusk and starts walking home through a softly lit residential area. Their music player happens to land on a track whose tempo aligns with their pace: the beat sits right on the rhythm of their steps, and the arrangement "opens up" just as they turn onto a quieter street lined with warm streetlights. Nothing in the system explicitly references the sunset or the neighborhood, yet the combination of a steady pace, mellow harmony, and restrained energy makes the moment feel authored, as if the city and the soundtrack are moving in the same meter.

Kari et al. [9] create such a meticulous adaptation of the music that the user currently listens to. By synchronizing musical elements with visual landmarks in the environment (such as tunnel exits or highway ramps), they produce an immersive and seamless integration of digital and physical components.

Contextual storytelling. On a weekend trip, someone drives through a city they don't know well, listening to an audiobook they selected days earlier. Mid-chapter, the narrator starts describing the very district the listener is currently passing through, mentioning a river-side promenade, a specific bridge silhouette, and the sound of trams in the distance. As the listener looks up and sees the same bridge



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI 2026 Workshop on Next Steps for Augmented Reality On-the-Move: Challenges & Opportunities, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).

and hears a tram pass, the story's setting and the physical place snap into alignment.

In my previously published research project *Story-Driven*, I implemented this scenario in a working prototype by utilizing on-demand AI-generated stories. Narrating an audiobook in real time allows for adjusting and guiding the plot in the right direction, no matter where the user is moving [2].

Visual AR retrospection. In the morning, a photo app shows a small "memory" card on the lock screen: a picture from exactly one year ago of friends sitting outside with coffee, wrapped in scarves. Today happens to be similarly cold and bright, and the user is about to meet one of those friends again. The resurfaced image is not just temporally correct; it also matches the day's atmosphere and upcoming social context, prompting the user to recall the mood, the conversation, and even the route they took.

Similarly, Dingler et al. [5] captured a continuous stream of images from the user's first-person perspective, which was later used to recall specific moments of the day, thereby augmenting the user's memory using contextual data. Nowadays, wearable AR devices can achieve the same effect on demand, providing an automatic, seamlessly integrated memory aid.

Several other examples exist as research prototypes: stress reduction through an ambient device [12] or breathing guide [6], real-time adjustment to audience engagement [7], and adaptive, immersive VR environments [11] are research areas that have been investigated, built, and evaluated by HCI researchers. While this highlights the potential of meticulous synchronization between the physical and real world, the publication dates also show how sparse the research coverage of such experiences is.

3 Discussion

This set of three different scenarios has already been implemented in existing research prototypes. Considering the technical possibilities that context-aware AR combined with generative AI provides, many more scenarios are feasible. When looking at these scenarios from a broader point of view, five key commonalities arise:

- **Anchor:** Every scenario has an initial set of variables that define the situational context. In the example of the synchronized music, this is the visual stimulus of the environment ("softly lit residential area") and the rhythm of the user's steps. Identifying and sensing the anchor is crucial for providing a well-synchronized experience.
- **Precision:** Depending on the type of user experience, elements may have to be adjusted more or less precisely, from multiple times per second to only once per hour. For example, in our storytelling project *Story-Driven*, we checked for new data every five seconds [2], whereas adaptive music playlisting only requires synchronization every couple of minutes [13].
- **Trigger:** Similar to the precision of adjustment, the type of trigger depends on the scenario: Sometimes, experience **boundaries** trigger a new experience cycle (e.g., a visual landmark appears [9]), in other cases, the adaptation process is a **continuous** one (e.g., *Story-Driven* [2]).
- **Interaction type:** Experiences can be synchronized **implicitly** based solely on contextual data. However, it can be

important to require **explicit** user input before adjusting the current user experience.

- **Stakes:** Some adjustments to a user experience can impose risks to the user, for example, through distraction or invasion of privacy. Particularly failure of synchronization can cause discomfort. Therefore, synchronized experiences require careful consideration of their relevance to the user and may be confirmed manually by the user (see *explicit interaction*).

In *Story-Driven*, we implemented a synchronized storytelling experience that adjusts to the user's current location. For example, if the user is walking towards a historic building or landmark, that place becomes the setting for the upcoming plot. Moreover, the story is synchronized at a fine level, so the place name is mentioned precisely when the user arrives [2]. In this case, the **anchor** is the point of interest that the user is heading towards, as well as the user's estimated time of arrival. The experience **precision** is medium-fine, since the story is adjusted every five seconds, but the time window for the user arriving at the landmark leaves room for error. Synchronization **triggers** in *Story-Driven* are continuous, since the user's estimated time of arrival is checked every five seconds, which triggers a story update. The **interaction type** is purely implicit; No user input is required. Finally, the **stakes** are quite low: if the synchronization fails, the story is still narrated, but at a different setting - no harm done to the user besides them being slightly bored.

Besides the main outcome of the user study, the design process of our research project revealed interesting insights for synchronized experiences. First, the anchor needs to be chosen carefully, as it determines the effectiveness of a synchronized experience. The mobile experience of *Story-Driven* would not have been transferable to other modalities if we had selected something other than the user's estimated time of arrival (e.g., speed would be significantly different for a car than for pedestrians). Second, the synchronization precision must be defined appropriately. Querying and processing contextual information every second might not be feasible in mobile scenarios with low computational power. Updating the context only every ten seconds might leave too much room for error, resulting in an out-of-sync experience. A suitable selection of boundary-based triggers combined with a feasible synchronization precision can result in a better experience than simple high-frequency update cycles. Finally, meticulously synchronized experiences are prone to fail and run out of sync - experience designers need to consider this and prepare for failure in consideration of the stakes.

Synchronized realities offer a large potential for immersive, frictionless, and "calm" experiences, both in visual and multimodal AR. It may take additional efforts to create a robust user experience that accounts for various contexts, but it also opens up the possibility for an almost "magical" experience. Context-aware, mobile AR offers the richest potential for magical experiences. In this reflection paper, I want to emphasize the importance of well-synchronized user experiences that "weave themselves into the fabric of everyday life until they are indistinguishable from it" [14].

4 Conclusion

References

- [1] Ronald T Azuma. 1997. A survey of augmented reality. *Presence: teleoperators & virtual environments* 6, 4 (1997), 355–385.
- [2] Jan Henry Belz, Lina Madlin Weilke, Anton Winter, Philipp Hallgarten, Enrico Rukzio, and Tobias Grosse-Puppendahl. 2024. Story-driven: exploring the impact of providing real-time context information on automated storytelling. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–15.
- [3] Jacob Bhattacharyya, Alessandro Vinciarelli, and Stephen Brewster. 2025. Birds of a Feather Augment Together: Exploring Sonic Links Between Real and Virtual Worlds in Audio Augmented Reality. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 1490–1500. doi:10.1109/ISMAR67309.2025.00153
- [4] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272.
- [5] Tilman Dingler, Passant El Agroudy, Huy Viet Le, Albrecht Schmidt, Evangelos Niforatos, Agon Bexheti, and Marc Langheinrich. 2016. Multimedia memory cues for augmenting human memory. *IEEE MultiMedia* 23, 2 (2016), 4–11.
- [6] Tan Gemicioglu, Thalia Viranda, Yiran Zhao, Olzhas Yessenbayev, Jatin Arora, Jane Wang, Pedro Lopes, Alexander T. Adams, and Tanzeem Choudhury. 2024. BreathePulse: Peripheral Guided Breathing via Implicit Airflow Cues for Information Work. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 188 (Nov. 2024), 33 pages. doi:10.1145/3702211
- [7] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A System for Implicit Audience Engagement Sensing Using Electroencephalography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (Denver, Colorado, USA) (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5114–5119. doi:10.1145/3025453.3025669
- [8] Jamil Hussain, Anees Ul Hassan, Hafiz Syed Muhammad Bilal, Rahman Ali, Muhammad Afzal, Shujaat Hussain, Jaehun Bang, Oresti Banos, and Sungyoung Lee. 2018. Model-based adaptive user interface based on context and user experience evaluation. *Journal on multimodal user interfaces* 12, 1 (2018), 1–16.
- [9] Mohamed Kari, Tobias Grosse-Puppendahl, Alexander Jagaciak, David Bethge, Reinhard Schütte, and Christian Holz. 2021. SoundsRide: Affordance-Synchronized Music Mixing for In-Car Audio Augmented Reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology (Virtual Event, USA) (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 118–133. doi:10.1145/3472749.3474739
- [10] Hayeon Kim and In-Kwon Lee. 2022. Studying the effects of congruence of auditory and visual stimuli on virtual reality experiences. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2080–2090.
- [11] Ilkka Kosunen, Mikko Salminen, Simo Järvelä, Antti Ruonala, Niklas Ravaja, and Giulio Jacucci. 2016. RelaWorld: Neuroadaptive and Immersive Virtual Reality Meditation System. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (Sonoma, California, USA) (IUI '16)*. Association for Computing Machinery, New York, NY, USA, 208–217. doi:10.1145/2856767.2856796
- [12] Diana MacLean, Asta Roseway, and Mary Czerwinski. 2013. MoodWings: a wearable biofeedback device for real-time stress intervention. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (Rhodes, Greece) (PETRA '13)*. Association for Computing Machinery, New York, NY, USA, Article 66, 8 pages. doi:10.1145/2504335.2504406
- [13] Xinxin Wang, David Rosenblum, and Ye Wang. 2012. Context-aware mobile music recommendation for daily activities. In *Proceedings of the 20th ACM international conference on Multimedia*. 99–108.
- [14] Mark Weiser. 1999. The computer for the 21st century. *ACM SIGMOBILE mobile computing and communications review* 3, 3 (1999), 3–11.
- [15] Jingyi Zhang, Jiaxing Huang, Sheng Jin, and Shijian Lu. 2024. Vision-language models for vision tasks: A survey. *IEEE transactions on pattern analysis and machine intelligence* 46, 8 (2024), 5625–5644.