Review: Real-Time Language Translator for Augmented Reality

RENATE ZHANG, Technical University of Vienna, Austria

Physical devices or human interpreters can help people to converse reliably in foreign languages. However, these approaches are obtrusive and distracting as they introduce a delay in the translation between the original and translated words, interrupting the natural flow of conversation. Real-time language translation in immersive environments could enable more seamless, face-to-face communication, removing any barriers and making conversations feel natural and uninterrupted. With the rise of large language models, research in language translation is rapidly developing, further pushing innovation for translating speech in augmented and virtual reality. Progress of this area is explored through literature research. Findings and implementations are examined, discussing the usability and limitations of the proposed systems across various applications and environmental contexts. Results include the insufficient depth in testing of the introduced systems and limited consideration for usability. Using audio for translated speech has been identified as a research gap, as well as investigating the social effects on users. Research of speech transcription can be seen as an example for displaying translation results as subtitles, in terms of interface recommendations and user study design.

CCS Concepts: • **Human-centered computing → Mixed / augmented reality**; **Collaborative interaction**.

Additional Key Words and Phrases: Augmented Reality, Virtual Reality, Real-time Translation, Automatic Speech Recognition, Large Language Models, Text-to-Speech Technology

ACM Reference Format:
Renate Zhang. 2024. Review: Real-Time Language Translator for Augmented Reality. 1, 1 (February 2024), 11 pages. https://doi.org/X

1 Introduction

Language barriers are still one of the biggest obstacles that hinder global communication [9]. People often rely on their smartphones to translate unknown languages or hire an interpreter in professional settings. However, having a device in between two parties speaking is obtrusive [16] and disrupts the natural flow of a conversation, even potentially disturbing the sense of privacy [9]. Although human interpreters can provide real-time translation through synchronous interpretation, where they speak while the original speaker is talking, this process is highly complex and demands interpreters to be skilled and stress resistant [7]. Effective interpretation also requires background knowledge in the respective fields, and interpreters often face intense pressure, which can lead to burnout [4]. Furthermore, the real voice of the speaker is masked by the voice of the translator. Meanwhile, the usage of a smartphone is often perceived as distracting by the user and gives the impression of a lack of interest towards others [16].

More recently, real-time transcription and translation of spoken language have become increasingly practical due to the developments in automatic speech recognition (ASR), large language models (LLMs) and text-to-speech (TTS) [5].

These technologies combined with immersive environment such as augmented reality (AR) or virtual reality (VR), could reduce the disadvantages in "traditional" translation and offer a more straightforward and less disruptive process [9]. Augmented and virtual reality glasses could lower the cognitive workload while improving comprehension and therefore providing a more engaging communication experience [2].

Author's Contact Information: Renate Zhang, renate.zhang@student.tuwien.ac.at, Technical University of Vienna, Vienna, Austria.

Language translation has been represented through audio translation using smart glasses with built-in speakers [1, 9] or depicted through subtitles, using AR displays [15] and VR glasses [2]. Similarly to depicting subtitles for a foreign language, various systems used continuous, real-time captions as visual representation for speech to aid deaf and hard-of-hearing people (DHH) using head-mounted displays (HMD) [10, 15–17].

In this review, the current research landscape of the real-time language is explored. The purpose of studies, their research design and findings are examined, and the overall quality assessed. Literature are compared, and the findings from these papers are combined to highlight the studies' shortcomings and differences. Overall the goal is to identify research gaps and discuss potential research perspectives. What are the challenges from existing publications, what can we learn and how can we use their findings in future studies?

2 Methodology

A literature review was carried out, analysing the usage of real-time translation and transcription technologies. The review was conducted in 2024 through a search on online databases using "SpringerLink", "IEEExplore", "ACM Digital Library", "ScienceDirect" and the Google Scholar search engine. The search covered the period from 2018 to 2024, focusing on publications from the last six years. The year 2018 was chosen due the release of several new AR and VR glasses, such as the Oculus Go, the Magic Leap One and HTC Vive Pro, to account for the recent advancements in AR and VR technologies.

The search was conducted using a combination of the following keywords and their abbreviations:"Real-time" AND "Language translation" AND "Augmented reality" OR "Virtual reality" AND "Glasses" OR "Display" OR "HMD". Only open-access resources or those accessible through the TU Vienna library and in publications written in English were considered. Additional relevant papers were identified by reviewing the related works sections of previously discovered studies, as well as using the "Similar papers" feature in Google Scholar and the listed databases. During the process, the term "Smart Glasses" was encountered in several studies and subsequently added to the list of keywords.

A thorough screening was applied to select articles relevant with the research focus. The papers were examined whether they involved translating speech and the usage of any form of augmented and virtual reality, not limited to glasses. Papers with focus on sign language interpretation and translation of written text where excluded as the technology relies on image and movement recognition instead of speech. Additionally, studies focusing solely on translation without any AR or VR element were not included, as well as studies with emphasis on language learning without a real-time translation component. While still relevant, these fields exceeds the scope of this paper and are therefore not considered.

In the first search phase, six relevant papers were found. However, during the first screening process it was noticed that most of the papers were lacking information on how the study was evaluated and had no clear findings. Under this circumstance, similar areas of research were considered, such as real-time captioning as visual representation. Displaying captions share similarities with generating translated subtitles in AR and VR, only without the language translation aspect. Consequently, the search term "Transcription" and "Captions", as well as their variations, were added, resulting in a higher number of qualitative publications. With the increased number of papers, a more selective approach was carried out, including only studies that conducted user studies and was cited more than 30 times. These additions contributed four, bringing the total to ten relevant papers. In future research this number will be further increased.

To organize the findings, studies were split broadly into two categories: real-time translation in AR and VR and real-time transcription in AR. Furthermore, a codebook was developed to extract relevant statements from each study and summarized on a Miro Board, see Appendix 5. Each section was coded in their respective colour: general purpose of the study (yellow), research method (blue), findings (green), shortcomings and improvements (pink), the darker colours post-its stand for higher relevancy and importance. Furthermore, certain criteria should be provided by studies in order to evaluate their quality. These included a description of a clear purpose, methodology design, explanation study's evaluation process and well-presented results.

3 Findings

The review of the research activity on real-time language translation and transcription using AR and VR led to the identification of ten studies, which are presented in Table 1 and Table 2. The authors, year of publication, purpose of study, research method and the findings of each paper are listed in the overviewing tables. In the following, a more detailed presentation of findings are presented.

3.1 Real-Time Translation and Captioning Technologies

Among all the papers, seven developed their own software to handle the translation and transcription mechanism [2, 9–11, 15–17], other methods included the usage of commercially available glasses with in-built translation [3], a paper prototype [19] and a purely literature-based study [18]. All but one paper focused on using subtitles to present translations, with one chose audio translation as the output [9].

Four out of the six studies regarding real-time translation focused on developing an own prototype with AR and VR devices [2, 9, 15, 19], while another study compared AR glasses with mobile apps [3] and one presented a review paper regarding AI-driven AR glasses for multilingual communication. The studies on real-time transcription for AR glasses explored different aspects of transcription, including user control of the caption interface [11], captioning in mobile [10] and all-day settings [16], as well as in group environments [17]. All studies involving transcription were designed for deaf and hard of hearing people.

3.2 Research Design

If an evaluation of a developed prototype was included the most common method was to conduct user studies [3, 10, 11, 16, 17]. Furthermore, to gather information beforehand, co-design sessions [16, 19], surveys [16] and formative studies [11] were used. Each study in Table 2 provided specific methodological information and conducted extensive user testing as well as aspects about the evaluations. However, only two papers [3, 19] regarding real-time translation in Table 1 provided further information on how their studies were evaluated and included the number of participants. The other three publications [2, 9, 15], which involved a prototype development, provided only rough information about the participant size and the procedure of the testing. For instance, it was mentioned that the prototype "was tested with a group of individuals with hearing challenges" [2], without any further elaboration .

These three papers showed signs of mediocre quality according to the defined quality criteria. The objectives and goals of these papers were too broadly defined, such as "aims to enhance communication for individuals facing hearing challenges and foster cultural connections" [2] or lacked further implications beyond the development of the AR glasses. For example, it was stated the goal was "to develop and test the use of AR audio glasses for translation between languages" [9], while another paper presented "a visual captioning interface to facilitate DHH individuals and connect people speaking different languages" [15]. These statement lacked specificity and did not include broader impacts or applications.

Summarizing, papers which included user studies concluded with more concise findings [3, 10, 11, 16, 17] while studies with no proper testing of the developed prototype have a lack of clear purpose [2, 9, 15]. Additionally, systems were tested in controlled environments with small, similar participant groups, limiting the applicability to diverse and real-world contexts [3, 16, 17].

3.3 User Experience

Systems like SpeechBubbles [17] and Wearable Subtitles [16] increase participation in both personal and professional settings by reducing cognitive load and enhancing comprehension [2]. Furthermore, wearable AR devices can create interpersonal connection and reduce feelings of isolation for DHH users [2, 11]. It was originally assumed that AR glasses would outperform mobile apps in various factors, such as usability and connectivity [3]. However, in direct comparison participants preferred the mobile app and found it more user-friendly and efficient. Reason for this, was the lack of accuracy and translation speed which impacted user experience. In fact, the AR glasses created higher mental loads and frustration. Without reliable translation quality, users cannot fully engage in conversation and benefit from the features of the AR glasses.

However, users welcomed customization of the subtitles and the transcription [2, 11, 15, 19], adjustable text size and font enhanced usability. Features like the spatial adjustment of captions allow users to follow conversations more naturally, thus reducing attention splits between speakers and visuals [10]. Therefore user-centric designs are recommended [18], which ease concerns over usability, but also privacy and societal acceptance. Real-time captions increase usability in mobile settings, allowing users to stay aware of their surroundings, e.g. while walking, compared to relying on phone-based services [11]. Other features like speaker identification and out-of-view indicators further help maintain focus and improve engagement in dynamic settings [10, 17].

3.4 Challenges

Linguistic nuances and cultural context is still a problem in translation systems in AR glasses [3, 9, 18]. Contextual errors in languages and non-native pronunciation can cause lower accuracy [9]. Systems with reliance of cloud processing for translation can introduce latency [18], which can, as mentioned before, impact user experience [3]. All hardware used were wireless systems which provides mobility, but are vulnerable on network availability issues [18],

Wearable designs, like lightweight glasses or subtle AR glasses, minimize stigma and ensure private captioning [16]. In contrast, if the device was too bulky and heavy, e.g. the HoloLens [10, 11], it would cause discomfort during extended use. Also, participants found wearing the HoloLens would draw attention which would make interactions feel unnatural for the user and the conversation partners. Furthermore, brightness issues in natural daylight and battery limitations reduce usability outdoors [11, 16]. Additionally, managing dynamic captioning settings during a conversation, such as repositioning text or handling overlapping speakers, can also overwhelm users, complicating their interaction with the system [11].

3.5 Design recommendations

Captions should be placed near the bottom-centre of the visual field in scrolling, line-by-line format to improve readability and reduce cognitive load [15, 17]. In general, text and visuals should be positioned close to the speaker, reducing visual attention split and therefore increases access to information [10, 15]. The speaker should be tracked, so captions should be automatic aligned with the speakers to reduce user effort [10]. In order to enhance mobility, lightweight and portable designs with broader language support are recommended for multilingual and mobile applications. In terms of

Table 1. Studies on real-time translation on AR and VR devices

Author/s	Year	Context and device	Purpose of study	Research method	Findings
Bal et al. [2]	2024	Speech Translation Subtitles VR Headset	Development of VR glasses with real-time speech translation through subtitles and captioning features.	Tested with a group of individuals with hearing challenges (n=?), evaluation only verbally stated.	NLP model results in high accuracy in real-time translation, increased understanding of participation of conversation,
Chen [3]	2023a	Speech Translation Subtitles AR Glasses	Comparison of AR glasses and apps for translation in academic and therapeutic settings.	User study (n=40) comparing AR glasses and mobile apps, quantitative measures and semi-structured interviews.	No quantitative differences; AR glasses show potential for natural interactions but face usability issues.
Hovde et al. [9]	2021	Speech Translation Translated Audio AR Glasses	Development of AR glasses with real-time auditory translation.	Tested on English, Spanish, and German for accuracy and speed (n=?).	Recognizing of context and tone, but cultural errors in German, lower accuracy for non-native speakers.
Li [15]	2023	Speech Translation Subtitles AR Display	Development of AR display equipped with real-time speech translation through subtitles and captioning features and summarize design space for AR interfaces	No formal testing of prototype.	Captions improve accessibility by aligning captions with speakers and reducing attention split.
Rasheed et al. [18]	2024	Speech Translation Subtitles and Translated Audio AR Glasses	Review of AI-driven AR glasses for real-time multilingual communication.	Analysed literature and case studies on translation accuracy and user challenges.	High potential for inclusivity with accurate translations but hindered by contextual, latency, and user adoption issues.
Simon et al. [19]	2024	Speech and Text Translation Subtitles AR Helmet Prototype	Development of AR glasses prototype with real-time speech and visual translation.	Four co-design sessions with participants (n=15) to refine features	Design recommendation for customizable subtitles and inclusion of cultural context.

Table 2. Studies on real-time transcription on AR glasses

Author/s	Year	Context and device	Purpose of study	Research method	Findings
Jain et al. [10]	2018a	Speech Captioning Captions AR Glasses	Development of AR glasses for DHH with real-time captioning and user control of caption interface.	Auto-ethnographic study using two prototypes tested in 10 academic settings over 6 weeks.	AR captions improved speaker interaction but faced usability issues like device discomfort, distractions and limited caption tracking.
Jain et al. [11]	2018b	Speech Captioning Captions AR Glasses	Examining the potential for mobile captions on HMDs and development of AR glasses for DHH with real-time captioning.	Formative study (n=12) identifying needs of deaf and hard of hearing people , and semi-controlled study (n=10) to evaluate prototype .	While walking captions can support communication access and improve attentional balance between speaker and environment.
Olwal et al. [16]	2020	Speech Captioning Captions AR Glasses	Development of 3D-printed AR glasses for DHH with real-time captioning in all-day scenario.	Large survey (n=501), pilot-study (formative in-lab research), study in-the-wild use, study in mobile and group conversations (n=24)	Light-weight prototype reduced social stigma but required enhancements in fit, text visibility, and transcription speed.
Peng et al. [17]	2018	Speech Captioning Speech Bubbles AR Glasses	Development of AR glasses for DHH with real-time captioning in group conversations.	Interviews (n=8), co-design sessions, and user studies (n=12) comparing traditional captions with the SpeechBubbles prototype.	Users preferred speech bubbles for associating speakers but noted issues with visual clutter and limited space. Improvements like multi-line bubbles and rising bubble animations.

participants, testing should be expanded to include diverse linguistic, cultural, and environmental contexts for more inclusive and robust systems [3, 19].

4 Discussion

Based on the findings above, the potential of real-time speech translation using AR glasses is still constrained by hardware limitations, suggesting the need of technological progress. The lack of user testing highlight the gap of developing practical and user-friendly solutions, while the usage of audio as a translation output has been neglected.

4.1 Current State of the Field

The field of real-time speech translation using AR glasses remains in an early stage of development. So far, the number of relevant papers are modest, as well as in regards of the academic quality. This can be explained through the technology itself, as AR glasses represent a new form of technology. In previous years the power of current augmented glasses were not sufficient enough to power language translation and not rely on external sources, e.g. connect to a mobile phone. Only in recent years, wearable AR glasses with language translation capabilities have been released to public, such as the XRAI Glass One [8], Magic Leap 2 [14], RayNeo X2 [6] and most notably the Ray-Ban Meta Glasses [13] and Snap's Spectacles [20]. Currently, only one scientific publication has used commercial available glasses for translation [3]. The current number of publications reflects the lengthy publication process. Excluding the research phase, peer review alone takes three to six months [12]. Since the models were released only recently, there has not been enough time to publish related papers.

However, research in speech captioning concerning DHH people produced a considerable number of publications. Specific fields, such as transcribing speech in group settings [11] or in mobile contexts [10] have already been investigated, which can serve as a template for creating translating language systems in similar environments. The development of speech translation using AR can align with approaches used for AR transcription, applying similar methodologies in prototype design and user studies.

4.2 Future Directions

So far, using subtitles for displaying speech translation is the preferred method. Captions for DHH have shown to improve the wearers involvement in conversation, this effect can be expected for subtitle translation in case of reliable latency and accuracy. However, other medium such as using audio have been neglected. Similar studies can be held, conducting research what kind of new problems using audio as an output will immerse. For instance, using voice-over for translation raises questions about how hearing both the original speech and its real-time translation simultaneously impacts comprehension and engagement during conversations. A more general question can then be formed, whether using subtitles would be a good option at all or if using auditory translation could yield better results. Determining the most effective method for presenting translated speech remains an open area for further investigation.

As ready-to-wear AR glasses become available, the discussion can shift from development to their social implications. Research can focus on how does wearing language translation glasses affect the wearer and the environment, instead on the hardware. Questions could be raised, such as in what way the glasses would influence the emotional connection between speakers of different languages. In the field of language learning, research could examine whether glasses foster or hinder improvement in learning a new foreign language. Further, similar to comparing AR translation glasses to mobile apps [3], comparison studies can be held with professional interpreters in terms of accuracy and user experience, possibly testify if AR glasses can substitute interpreters.

Publications of speech translation show a deficit in user testing. The field of speech transcription can be used as a guide, already existing findings can be considered when designing for language translation interfaces. For instance, the display and mechanics of the subtitles is highly comparable to captions for DHH, thus existing design recommendations can be reapplied when incorporating language translation subtitles. Furthermore, user-centric design is recommended, as conducting formative studies beforehand can lead to more user-friendly outcomes. As cultural context is still an issue in language translation, future research should prioritize development of language algorithms and creation of efficient hardware to improve translation speed and precision.

In future studies, a broader spectrum of papers will be considered and a higher number of analysed paper included. The citation count will be reconsidered as a metric to ensure the inclusion of the most recent publications.

5 Conclusion

This study presents a literature review of research on real-time language translation using speech, which includes the analysis of ten papers. While there is progress in developing real-time language translation systems using AR and VR, several challenges and limitations remain. Many studies lack evaluation methodologies and clear research objectives and findings. However, systems focusing on transcription for DHH can serve as an example for research and prototype design, e.g. including user studies in research to enhance usability.

While current systems increase participation and improving understanding, AR glasses still struggle with linguistic nuances while facing problems with latency. User satisfaction can be improved by providing customization and keeping to design recommendations, such as placing the text near the speaker and on the bottom-centre in scrolling, line-by-line format. The device in general should be light-weight and subtle to increase long-time wear and reduce social acceptance. Furthermore, research on speech translation in AR and VR using audio as output method has been identified as a promising research gap where further research on social effects is needed.

Real-time language translation in immersive environments is a growing trend, that represents a dynamic and fast developing field. This review is well-timed, as new AR glasses are expected to be released on the market and subsequently come in contact with the wide public. Researchers may benefit from the findings and results from this review as the existing limitations and gaps are provided.

References

[1] Lovitt Andrew, Miller Antonio John, Robinson Philip, and Selfon Scott. 2021. Natural language translation in augmented reality (AR). https://patents.google.com/patent/US11068668B2/en

[2] Advait Bal, P. Anjana, K. T. Gayathri Devi, Vishaal S. Upot, S. Abhishek, and T. Anjali. 2024. *AR Glasses for Real-Time Translation and Subtitling*. 279–288. https://doi.org/10.1007/978-981-97-6678-9_25

[3] Shenghao Chen. 2023. EYES ON THE FUTURE: A MIXED-METHODS EVALUATION OF REAL-TIME TRANSLATION IN AUGMENTED REALITY GLASSES. (2023).

[4] Tian Chen. 2023. The interplay between psychological well-being, stress, and burnout: Implications for translators and interpreters. *Heliyon* 9, 8 (Aug. 2023), e18589. https://doi.org/10.1016/j.heliyon.2023.e18589

[5] Wenqian Cui, Dianzhi Yu, Xiaoqi Jiao, Ziqiao Meng, Guangyan Zhang, Qichao Wang, Yiwen Guo, and Irwin King. 2024. Recent Advances in Speech Language Models: A Survey. (10 2024). http://arxiv.org/abs/2410.03751

[6] Heaney David. 2024. RayNeo X2 Review: Are The First Standalone AR Glasses More Than A Tech Demo? (Sept. 2024). https://www.uploadvr.com/rayneo-x2-standalone-ar-glasses-review/ Publication Title: UploadVR.

[7] Z.Z. Gafarova, M.A. Bozorova, Sh.Sh. Jumayeva, L.I. Idiyeva, and L.U. Radjabova. 2020. Synchronous Translation – a Complex Set of Cognitive Processes. *International Journal of Psychosocial Rehabilitation* 24, 1 (Jan. 2020), 403–407. https://doi.org/10.37200/IJPR/V24I1/PR200143

[8] XRAI Glass. 2024. (2024). https://xrai.glass/blog/introducing-the-xrai-ar-one/

[9] Ian N. Hovde, Forrest S. Kelley, Ryan J. Kearney, and Douglas E. Dow. 2021. Aural Language Translation with Augmented Reality Glasses. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, Vol. 403 LNICST. Springer Science and Business Media Deutschland GmbH, 61–70. https://doi.org/10.1007/978-3-030-92163-7_6

[10] Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. 2018. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive*

Systems. ACM, Hong Kong China, 7–11. https://doi.org/10.1145/3197391.3205404

[11] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People who are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3234695.3236362

[12] Springer Nature Journal. 2024. (2024). https://support.springer.com/en/support/solutions/articles/6000131708-timescale-to-publish-an-article-for-a-springer-nature-journal

[13] Knibbs Kate. 2024. I Wore Meta Ray-Bans in Montreal to Test Their AI Translation Skills. It Did Not Go Well. (June 2024). https://www.wired.com/story/meta-ray-ban-ai-translation-skills-do-not-work-well/ Publication Title: Wired.

[14] Magic Leap. 2022. (2022). https://www.magicleap.com/magic-leap-2

[15] Jingya Li. 2023. Augmented Reality Visual-Captions: Enhancing Captioning Experience for Real-Time Conversations. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 14037 LNCS. Springer Science and Business Media Deutschland GmbH, 380–396. https://doi.org/10.1007/978-3-031-34609-5_28 ISSN: 16113349.

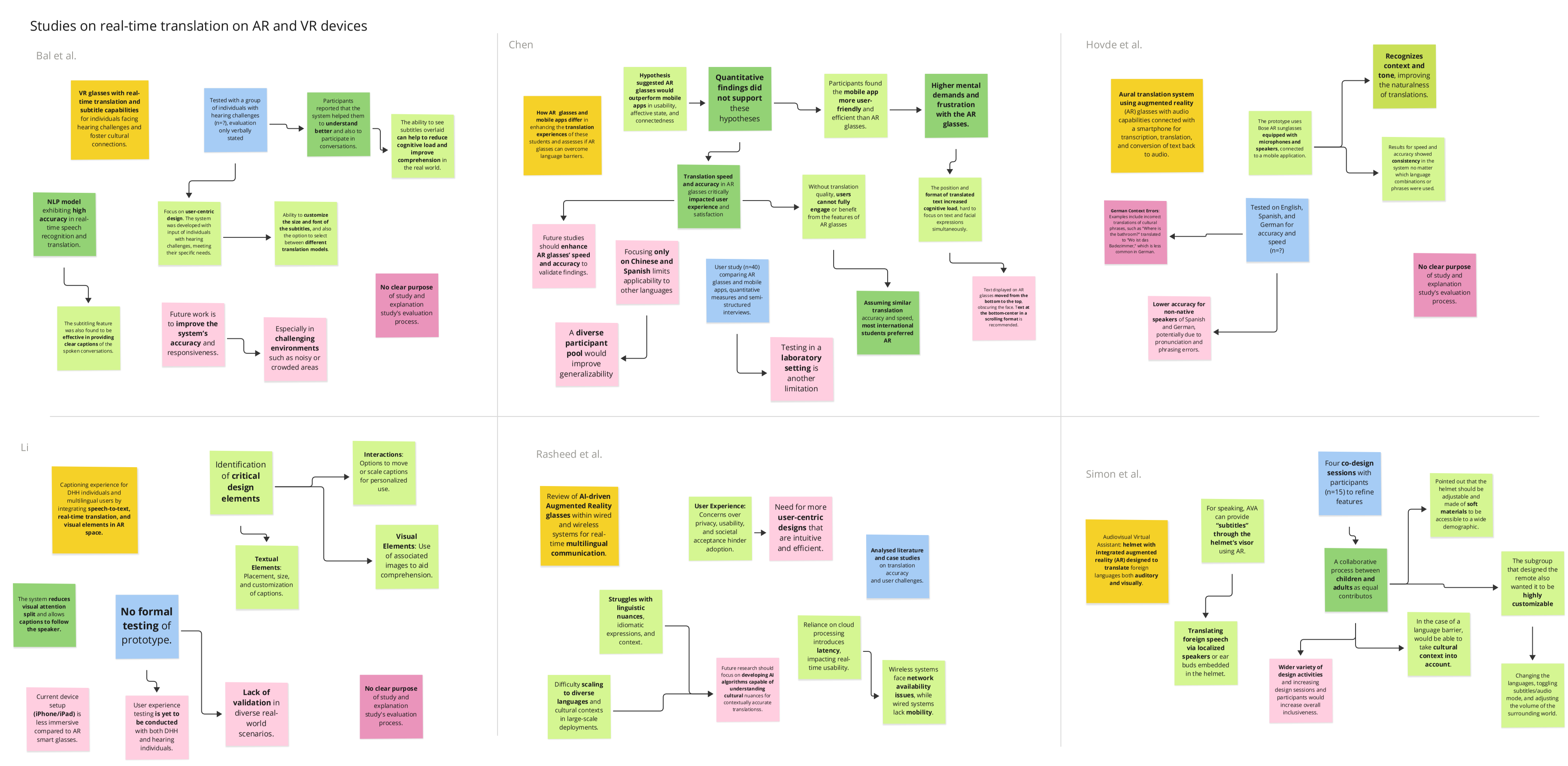
[16] Alex Olwal, Kevin Balke, Dmitrii Votintcev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. 2020. Wearable subtitles: Augmenting spoken communication with lightweight eyewear for all-day captioning. In *UIST 2020 - Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. Google Research, 1108–1120. https://doi.org/10.1145/3379337.3415817

[17] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–10. https://doi.org/10.1145/3173574.3173867

[18] Zainab Rasheed, Sameh Ghwanmeh, and Aziz Almahadin. 2024. A Critical Review of AI-Driven AR Glasses for Realtime Multilingual Communication. In *2024 Advances in Science and Engineering Technology International Conferences (ASET)*. IEEE, Abu Dhabi, United Arab Emirates, 1–6. https://doi.org/10.1109/ASET60340.2024.10708749

[19] Tori Simon, Upama Thapa Magar, Shrutee Dwa, Michael S Wendell, Benjamin J Bettencourt, and Jerry Alan Fails. 2024. The Audiovisual Virtual Assistant (AVA): The Augmented Reality Translation Helmet that Breaks Language Barriers. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*. ACM, Delft Netherlands, 1008–1011. https://doi.org/10.1145/3628516.3661604

[20] Spectacles. 2024. (2024). https://newsroom.snap.com/sps-2024-spectacles-snapos?lang=en-US



Studies on real-time translation on AR and VR devices
Bal et al.
VR glasses with real-time translation and subtitle capabilities for individuals facing hearing challenges and foster cultural connections.
Tested with a group of individuals with hearing challenges (n=?), evaluation only verbally stated
Participants reported that the system helped them to understand better and also to participate in conversations.
The ability to see subtitles overlaid can help to reduce cognitive load and improve comprehension in the real world.
NLP model exhibiting high accuracy in real-time speech recognition and translation.
Focus on user-centric design. The system was developed with input of individuals with hearing challenges, meeting their specific needs.
Ability to customize the size and font of the subtitles, and also the option to select between different translation models.
The subtitling feature was also found to be effective in providing clear captions of the spoken conversations.
Future work is to improve the system's accuracy and responsiveness.
Especially in challenging environments such as noisy or crowded areas
No clear purpose of study and explanation study's evaluation process.
Chen
How AR glasses and mobile apps differ in enhancing the translation experiences of these students and assesses if AR glasses can overcome language barriers.
Hypothesis suggested AR glasses would outperform mobile apps in usability, affective state, and connectedness
Quantitative findings did not support these hypotheses
Participants found the mobile app more user-friendly and efficient than AR glasses.
Higher mental demands and frustration with the AR glasses.
Translation speed and accuracy in AR glasses critically impacted user experience and satisfaction
Without translation quality, users cannot fully engage or benefit from the features of AR glasses
The position and format of translated text increased cognitive load, hard to focus on text and facial expressions simultaneously.
Future studies should enhance AR glasses' speed and accuracy to validate findings.
Focusing only on Chinese and Spanish limits applicability to other languages
User study (n=40) comparing AR glasses and mobile apps, quantitative measures and semi-structured interviews.
Assuming similar translation accuracy and speed, most international students preferred AR
Text displayed on AR glasses moved from the bottom to the top, obscuring the face. Text at the bottom-center in a scrolling format is recommended.
A diverse participant pool would improve generalizability
Testing in a laboratory setting is another limitation
Hovde et al.
Aural translation system using augmented reality (AR) glasses with audio capabilities connected with a smartphone for transcription, translation, and conversion of text back to audio.
The prototype uses Bose AR sunglasses equipped with microphones and speakers, connected to a mobile application.
Recognizes context and tone, improving the naturalness of translations.
Results for speed and accuracy showed consistency in the system no matter which language combinations or phrases were used.
German Context Errors: Examples include incorrect translations of cultural phrases, such as "Where is the bathroom?" translated to "Wo ist das Badezimmer," which is less common in German.
Tested on English, Spanish, and German for accuracy and speed (n=?)
No clear purpose of study and explanation study's evaluation process.
Lower accuracy for non-native speakers of Spanish and German, potentially due to pronunciation and phrasing errors.
Li
Captioning experience for DHH individuals and multilingual users by integrating speech-to-text, real-time translation, and visual elements in AR space.
Identification of critical design elements
Interactions: Options to move or scale captions for personalized use.
Textual Elements: Placement, size, and customization of captions.
Visual Elements: Use of associated images to aid comprehension.
The system reduces visual attention split and allows captions to follow the speaker.
No formal testing of prototype.
Current device setup (iPhone/iPad) is less immersive compared to AR smart glasses.
User experience testing is yet to be conducted with both DHH and hearing individuals.
Lack of validation in diverse real-world scenarios.
No clear purpose of study and explanation study's evaluation process.
Rasheed et al.
Review of AI-driven Augmented Reality glasses within wired and wireless systems for real-time multilingual communication.
User Experience: Concerns over privacy, usability, and societal acceptance hinder adoption.
Need for more user-centric designs that are intuitive and efficient.
Analysed literature and case studies on translation accuracy and user challenges.
Struggles with linguistic nuances, idiomatic expressions, and context.
Difficulty scaling to diverse languages and cultural contexts in large-scale deployments.
Future research should focus on developing AI algorithms capable of understanding cultural nuances for contextually accurate translationss.
Reliance on cloud processing introduces latency, impacting real-time usability.
Wireless systems face network availability issues, while wired systems lack mobility.
Simon et al.
Audiovisual Virtual Assistant: helmet with integrated augmented reality (AR) designed to translate foreign languages both auditory and visually.
For speaking, AVA can provide "subtitles" through the helmet's visor using AR.
Translating foreign speech via localized speakers or ear buds embedded in the helmet.
Four co-design sessions with participants (n=15) to refine features
A collaborative process between children and adults as equal contributos
Pointed out that the helmet should be adjustable and made of soft materials to be accessible to a wide demographic.
The subgroup that designed the remote also wanted it to be highly customizable
Changing the languages, toggling subtitles/audio mode, and adjusting the volume of the surrounding world.
In the case of a language barrier, would be able to take cultural context into account.
Wider variety of design activities and increasing design sessions and participants would increase overall inclusiveness.

Studies on real-time transcription on AR glasses

Received 03 November 2024; revised 17 November 2024; accepted 12 January 2025