Enhancing the Audience Experience for VR and AR Theatre with AI-generated Subtitles

Irene Viola
Centrum Wiskunde & Informatica
Amsterdam, Netherlands
irene@cwi.nl

Moonisa Ahsan
Centrum Wiskunde & Informatica
Amsterdam, Netherlands
moonisa@cwi.nl

Olga Chatzifoti
NKUA
Athens, Greece
Maggioli Group
Athens, Greece
olchatz@di.uoa.gr

Atanas Yonkov
Centrum Wiskunde & Informatica
Amsterdam, Netherlands
atanas.yonkov@cwi.nl

Eleni Oikonomou
Athens-Epidaurus Festival
Athens, Greece
e.oikonomou@aefestival.gr

Ioannis Radin
Athens-Epidaurus Festival
Athens, Greece
i.radin@aefestival.gr

PawełMąka
Maastricht University
Maastricht, Netherlands
pawel.maka@maastrichtuniversity.nl

Abderrahmane Issam
Maastricht University
Maastricht, Netherlands
abderrahmane.issam@maastrichtuniversity.nl

Pablo Cesar
Centrum Wiskunde & Informatica
Amsterdam, Netherlands
TU Delft
Delft, Netherlands
p.s.cesar@cwi.nl

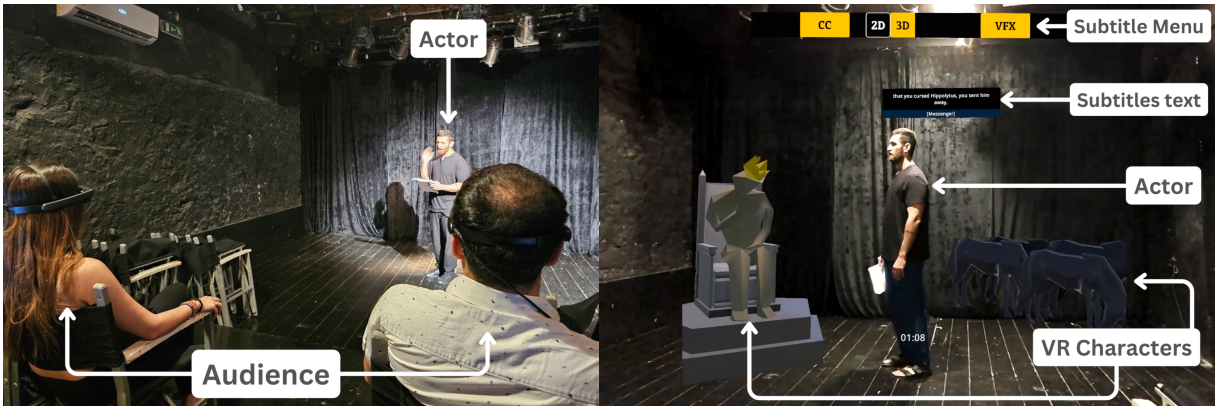

Figure 1: AR Greek Theatre Play Setup: (Left) Audience view from third-person perspective (Right) Immersive view from the headset

Abstract

Recent technological developments on AI and immersive media are transforming the artistic landscape, providing novel mechanisms for artists and audiences. Following a human-centric approach, together with a theatre company in Greece, this paper investigates how subtitle placement affects user experience and cognitive load in a live theatre performance enhanced by AR glasses. To do so, we design and develop a system for displaying subtitles in VR and AR. We evaluated the system in two conditions ($N = 19; N = 12$), both in a controlled environment (VR) and an actual theatre (AR). In the latter, we integrate AI solutions to provide automatic captioning and translation in real time, and VFX to further augment the experience. Our quantitative and qualitative results showed no difference between subtitle placements in terms of cognitive load and user experience, with users equally liking the two proposed approaches. Results also highlighted the perceived usefulness of AR to enhance theatre performances, indicating new paths for wider accessibility and further immersion.

VRST '25, Montreal, QC, Canada

ACM ISBN 979-8-4007-2118-2/25/11
https://doi.org/10.1145/3756884.3766057

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; **Virtual reality**; **User studies**.

Keywords

User Experience Design, Virtual Reality, Augmented Reality, Quantitative Methods, Usability Study

ACM Reference Format:

Irene Viola, Moonisa Ahsan, Olga Chatzifoti, Atanas Yonkov, Eleni Oikonomou, Ioannis Radin, PawełMąka, Abderrahmane Issam, and Pablo Cesar. 2025. Enhancing the Audience Experience for VR and AR Theatre with AI-generated Subtitles. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3756884.3766057

1 Introduction

The evolution in media accessibility has enhanced features like subtitles [12]. Subtitles are critical in furthering comprehension and inclusivity across audiovisual media, bridging linguistic, auditory, and cognitive barriers for diverse audiences. Subtitles have evolved throughout history, with studies showing their positive impact and sometimes even necessity on viewer comprehension of the media [17]. Research indicates that bilingual subtitles, combining intra-lingual and interlingual elements, provide linguistic support, making videos easier to comprehend and aiding learning [4]. Additionally, broadcasts with subtitled original versions have been linked to improved English proficiency worldwide, especially in listening comprehension skills [39]. Subtitles play an important role in making video content accessible [25] to diverse populations of viewers, such as non-native language speakers, the hard of hearing, and people with learning disabilities. They enable such viewers to understand the video content.

The advent of eXtended Reality (XR) opens new avenues for subtitling beyond traditional video content. Similar to subtitles in two-dimensional spaces, subtitles in XR provide a means for understanding audiovisual content in a three-dimensional space. One such example is attending a theatrical play in a foreign language [43], [31]. However, there is a lack of guidelines on implementing subtitles or live captioning in XR and how it differs from traditional television captioning [7]. Studies have been conducted to understand user preferences for different Virtual Reality (VR) captioning styles [6]. Generally, these studies tested different caption movement behaviours, such as head locked, lag, and appear captions, while participants watched live-captioned presentations in VR [6], [30]. However, such studies were only conducted in VR settings, and the impact of subtitle placement on Augmented Reality (AR) scenarios, where the performance is happening live in front of the audience, has never been tested before.

In this paper, we aim to study how subtitle placement affects user experience in a live XR theatre scenario. Based on the requirements collected by Lee et al. [23], we design an application that can support different subtitle placements: 2D, that is, tethered to the user's view, and 3D, that is, anchored to the world view. To ensure that our approach is human-centred and rooted in the actual needs of theatre companies, we designed our application together with a theatre company based in Greece. To allow for rigorous evaluation of the subtitle placements, we first perform a simulation of the live theatre scenario in a VR environment, with virtual actors and props. We conducted an initial study with 19 participants to assess which subtitle placement is optimal, using mixed methods. Based on the findings of the initial study, when then performed a study with a live theatre performance, augmented through the use of AR glasses with subtitles. To aid in the task, we designed and deployed two Machine Learning (ML) solutions for automated captioning and translation. Whereas the use of ML algorithms for live performance has been considered before [20], their evaluation has only been based on objective metrics. Conversely, we aim to evaluate our approach by performing live capturing and translation in front of an audience. Unlike traditional subtitling methods that require significant manual effort, our system automates the process, making it more efficient and scalable. Moreover, as AR allows the superimposition of virtual elements with real-world scenes, new avenues arise in augmenting the play with computer-generated Visual Effects (VFX) [37]. Thus, in our study, we also explored how VFX can enhance the user experience.

The contributions of the paper can be summarised as follows:

- We design and evaluate an application for AR theatre that allows for overlaying subtitles and VFX on top of a live performance. We base our design on user-centric requirements to ensure the application meets the needs of audiences and theatre organizers.
- We integrate ML solutions to perform automatic captioning and translation in real time, allowing for minimal human intervention and wider accessibility to the play in languages not serviced by official translations.
- We evaluate which subtitle placement style (2D or 3D) leads to better user experience and lower cognitive load. To do so, we perform two experiments: the first ($N = 19$) in a simulated VR theatre, which allows us to control and reproduce all conditions; the second ($N = 12$) during a live performance enhanced with AR glasses. We use mixed methods to assess the experience, drawing insights from questionnaires and interviews. No differences were observed between the two placements, in terms of cognitive load and user experience, as confirmed by user interviews. Moreover, our results highlight the potential of AR for making theatre performances more accessible, while remarking on the technical limitations that still hinder their widespread adoption.

2 Related work

2.1 VR and AR for theatre

The use of immersive technologies for enhancing theatre experiences has long been discussed in the literature [15, 29], and has further exploded in popularity after the COVID-19 pandemic prevented traditional access to live venues [2]. Pike explores the possibilities for XR technologies for theatre, noting that "live performing arts combine different media elements and support interaction and connection between the audience and the stage. In an historical sense, artists have always used impressive immersion tricks, such as physical surroundings with fake perspective [...] to immerse the viewer in illusory reality" [21], placing the audience as the main body that would decree the success or failure of the technology [32]. Ferreira et al. consider the challenges related to create effective AR experiences for theatre, and propose an authoring application to support AR performances, for example by adding visual effects [13]. Other studies focus on using immersive technologies to improve accessibility for the hearing impaired [19, 28] and the elderly [27], and considering integrating AI tools for facilitating the captioning

and translation [20]. In our work, we use AR to augment the theatre experience with live captioning and translation, considering the further possibilities that AR can offer through the addition of VFX.

2.2 Subtitles in XR

Studies on subtitle placement for 2D screens indicated that modes of presentation that follow the speaker, as opposed to being placed at the bottom of the screen, can lead to lower cognitive load and better comprehension [9, 22]. How such results translate to XR settings has been preliminary studied for omnidirectional content, which offers 3 degrees of freedom. Rothe et al. [38] compared two main subtitle variants using HMDs: static (tethered to the head's movement, or 2D), and dynamic (anchored on the virtual world, or 3D, and close to the speaking person). Results indicated a preference for dynamic due to higher presence scores, fewer sickness symptoms, and lower workload. Given the prevalence of virtual reality sickness among viewers, the behaviour of subtitles should not exacerbate this issue, further complicating their design and placement. Subtitles must be situated intuitively for viewers, ensuring they don't require excessive conscious effort to locate. Brown et al. evaluated the user experience of four types of subtitle placements: evenly spaced in the sphere, following the head immediately, follow with lag, and appear in front, then fixed (2D [1]). To perform the evaluation, they adapted the UX framework for subtitles [10]. They found that the general preference towards following the head immediately is better than other behaviours regarding ease of locating subtitles. However, the study is limited to focusing only on subtitle behaviour and did not extensively explore other design aspects that could impact user experiences, like font size or colour. Brescia-Zapata et al.[6] explored creative subtitle production in immersive environments, particularly focusing on eye-tracking to test subtitle placement in VR. The collected eye movement data results indicate that the focus group supports the preference for subtitles fixed relative to the speaker or overhead-locked subtitles (dynamic, or 3D, subtitles). Rzayev et al. [41] investigated the impact of text position and reading mode on comprehension and workload in AR while walking and sitting. Participants reported that reading from the centre or bottom-centre position resembled reading from a computer screen or book. In contrast, continuous reading from the top-right position led to eye strain. Further studies indicated that displaying translations in different positions on AR glasses could also impact comprehension and task load [40].

All the previous work has either focused on 360-degree videos [1, 38], or focused on studying how word placement in AR for tasks as language learning could impact the user experience and task load [40, 41]. However, no study has been conducted to evaluate which subtitle placement would be optimal in the case of a live theatre performance enhanced by AR glasses. In our study, we first test the subtitle placement by mimicking the AR scenario in a VR environment; then, we further assess it in an actual AR theatre scenario.

3 Testing VR subtitle placement

3.1 Experiment design and technical setup

To effectively test which subtitle placement would be optimal for a VR theatre scenario, we recreated a theatre play in Unity. In particular, an excerpt from the Hippolytus tragedy by Euripides was chosen to be displayed in the original Ancient Greek language. The excerpt was four minutes and thirty second long, and consisted of a monologue. The scene was recreated in Unity using low fidelity polygons, and consisted of a main character ("messenger") standing on stage delivering the monologue, a secondary character ("king") sitting on a throne, alongside several visual effects (VFX) that would follow the monologue as it unfolded (e.g., horses would appear as they would be mentioned by the messenger). An official translation to English made by a translator was used for the subtitles. English was the only language offered.

Figure 2: VR Application interface with 2D and 3D caption

Two caption positioning modes were selected for evaluation: one called 2D (fig. 2 right), where the caption frame was tethered to the HMD and followed its movements without delay, and one called 3D (fig. 2 left), where the caption frame was tethered to the physical theatrical stage, following common placements in the literature in terms of VR and AR subtitling [6, 38]. Users were given extensive customization options in both modes to adapt the display to their needs for improved readability. The customization options in both modes were identical. In absence of accessibility standard for immersive environments, web-based accessibility standards influenced (Web Content Accessibility Guidelines (WCAG) 2.0) the design of the customization options, which included: font size (in increments of 1 point), text position on the vertical axis (in increments of 10% of the view frame's height in pixels), text background contrast (as a slider with values 0-1 which set the opacity of a black frame), overall brightness of the screen (as a slider of 0-1) and most importantly, viewing distance (as a slider with a minimum value of 0.70cm and a maximum of 350cm). The entire user interface panel could be toggled on/off using a button on the XR device's controller.

The application was developed in Unity v2022.3.22f. A Windows 10 computer with an Intel i7-8700K processor, NVIDIA GeForce GTX1080 graphics card,and 32 GB of RAM was used to run the VR experience. Using a QuestLink wireless connection, the output was displayed on an untethered Meta Quest Pro device.

3.2 Experiment methodology and planning

3.2.1 Experiment methodology. For the study, a mixed-method methodology was chosen, comprised of quantitative questionnaires and semi-structured interviews. For the quantitative part, several factors were selected to study how subtitle placement influences the user experience. In particular, we tested for *Cognitive load* using the NASA TLX questionnaire [16], *Usability* using the System Usability Scale (SUS) [8], *User Experience (UX)* adapting questions from the UX for Subtitles framework [1], *Behavioural Intention* [45], and *Quality of Experience (QoE).* Additionally, we tested for typical VR-related factors, such as *Immersion* [18] and *Presence* [26]. To account for cybersickness, we administered the Simulator Sickness Questionnaire (SSQ) before and after the experiment, following guidelines from the literature. The full questionnaires administered

to the users are made available as supplemental material. For the qualitative part, we based our questions on Brown's white-paper regarding dynamic subtitles in 360-degree environments[1].

3.2.2 Participants. We conducted a power analysis with the software program G*Power [11], with the goal of obtaining 80% power to detect a large effect size of 0.8 at the standard 0.05 alpha error probability using a Wilcoxon two-tailed test, leading to a minimum participant pool of 15 subjects. The experiment included 19 participants (7 female, 12 male, mean age: 35 [min 22, max 71]). On a scale from one to five (best), the mean self-rated experience with immersive technologies was 2.8 (SD=1.22, median=3). Preferentially, participants were theatre-goers, but the only strict criterion was to be a non-native Greek speaker. The mean self-rated attendance in performances where string values were converted to numbers 1 (less than once a year) to 4 (more than 3 times a year) was 2.29 (SD=1.19, median=2). Recruitment aimed to ensure diversity in age, gender, and XR familiarity to remove potential sampling bias. Post-hoc achieved power for the required effect size was 89%.

3.2.3 Procedure. The study followed a within-subject methodology; thus, all subjects experienced both the 2D and 3D variants of the subtitles, in randomized order. The same excerpt was used in all the sessions. At the start of the experiment, the researchers explained the purpose of the study, after which the participants had to sign a consent form. Demographic data was also collected, and the SSQ was given. Then, the study proceeded as follows:

- **Step 0: tutorial and intro.** The participants were instructed to sit on a swivel chair and wear the Meta Quest Pro. They entered the VR environment and were given a few minutes to explore the environment to cultivate familiarity and mitigate potential discomfort during subsequent steps in the study. They were recommended to look around freely for a few minutes until they got accustomed to the VR environment and controllers; the duration varied between participants.
- **Step 1: first theatre session.** Participants experienced the theatre play with either 2D or 3D subtitle placement. The starting placement was randomized to avoid confounding factors. Their task was simply to watch the play and adjust the subtitles if they felt the need. After the experience, they were asked to fill the SSQ, NASA-TLX, SUS, UX, iPQ, Immersion, Behavioural Intention, and QoE questionnaires. A break of 10 minutes was then enforced to avoid fatigue.
- **Step 2: second theatre session.** Before the session, the SSQ was administered again to see whether the break led to lowering in cybersickness symptoms. Participants experienced the theatre play with either 3D or 2D subtitle placement, depending on which placement they had experienced in the previous step. After the experience, they were asked to fill the SSQ, NASA-TLX, SUS, UX, iPQ, Immersion, Behavioural Intention, and QoE questionnaires. Following an optional short break (0-15 minutes), participants engaged in a semi-structured interview. The interview was audio-recorded. Upon conclusion, participants received debriefing information and were granted a voucher for their participation.

3.2.4 Data analysis. For each questionnaire and subscale, the mean μ, standard deviation σ, median M and interquartile range IQR are shown separately for the 2D and 3D case. The quantitative data was tested for normality using the Shapiro-Wilk test (p values for each item are in supplemental material). If data was normally distributed, the Welch's t-test was administered. If normal, Welch's t-test with unequal variances is applied on the data, reported in terms of p-value, t statistic, Cohen's d, and degrees of freedom; otherwise, the results of the Wilcoxon signed-rank test are shown in terms of p-value, z-value, and effect size r. All questionnaires were grouped into the corresponding items based on the original papers. The interviews conducted for this study were analyzed using the thematic analysis approach as outlined by Braun and Clarke [5]. Two coders were used to condense the results into themes. Individual participants are labeled P1-P19. The number of participants who agreed with the given statement is indicated in parentheses. Please note that in the quotes, "2D" and "static" are used interchangeably; same for "3D" and "dynamic".

3.3 Results

3.3.1 Quantitative results. Table 1 showcases the results of the questionnaires administered to the participants. Results show favourable results in terms of usability for both subtitle placements (SUS: $\mu = 83.42$ vs $\mu = 83.82$ for the 2D and 3D case, respectively, on a scale from 0 to 100), corresponding to the second quartile in both cases [3]. Results also indicate that the subtitle placement, either in 2D or 3D, led to low cognitive load, as all the items in the NASA-TLX scale have low to medium values. In terms of UX, results showed a favourable performance for both 2D and 3D subtitle placements ($\mu = 5.14$ vs $\mu = 4.97$ for the 2D and 3D case, respectively, in a 7-item Likert scale). Similarly, behavioural intention indicated very favourable results for both cases ($\mu = 4.65$ vs $\mu = 4.74$ for the 2D and 3D case, respectively, in a 5-point Likert scale). The QoE questions in term of interface and general experience also indicated favourable perception ($\mu = 3.90$ and $\mu = 3.84$ vs $\mu = 4.05$ and $\mu = 3.84$ for the 2D and 3D case, respectively, in a 5-point Likert scale). No statistical differences were observed for any of the items.

In terms of immersion, both subtitle placements scored high on Person-VE interaction ($\mu = 4.67$ vs $\mu = 4.18$ for the 2D and 3D case, respectively, in a 7-item Likert scale), and on the Immersion subscale, albeit slightly lower (($\mu = 3.79$ vs $\mu = 3.76$ for the 2D and 3D case, respectively). In terms of presence, high values were observed for subscales presence, spatial presence, and involvement ($\mu = 4.74$, $\mu = 4.23$ and $\mu = 4.46$ vs $\mu = 4.26$, $\mu = 4.19$ and $\mu = 4.30$ for the 2D and 3D case, respectively, in a 7-item Likert scale), whereas lower scores were given to the subscale related to realism ($\mu = 2.99$ vs $\mu = 2.92$ for the 2D and 3D case, respectively). This is easily explained considering that our environment consisted of low-fidelity, low-polygon-count objects, as realism was not one of the objectives of our experiment. In terms of statistical analysis, we observed statistical differences only for the subscale *Presence* of the IPQ, with a medium effect size ($p = 0.02, z = 2.26, r = 0.37$).

The SSQ results showed a no significant increase of symptoms after the sessions, indicating that the play did not induce cybersickness. Complete results can be seen in the supplemental material.

3.3.2 Qualitative results. Overall, opinion was good (15) - (*P12: "...The theatre experience was quite interesting. I felt like I was in the film or the scene. It was extremely good..."*) Regarding the interface,

Table 1: Results of the questionnaires administered for the VR theatre experiment, for 2D and 3D subtitle placement, along with results of the Welch's t-test (upper part) or Wilcoxon signed-rank test (lower part) comparing the two, depending on the results of the normality test. In bold we indicate statistically significant differences.

		2D				3D				Comparison			
		μ	σ	M	IQR	μ	σ	M	IQR	p	t	d	df
SUS		83.42	10.15	85	14.38	83.82	8.22	85	12.50	0.90	-0.13	0.04	34.52
UX for subtitles		5.14	1.23	3.14	0.50	4.97	1.18	3	0.43	0.68	0.42	0.13	35.92
Immersion	Person-VE interaction	4.67	1.7	5	2.5	4.18	1.6	4	2.17	0.37	0.92	0.29	35.86
Presence	Involvement	4.46	1.82	4.5	3.12	4.30	1.89	4.5	3.44	0.80	0.26	0.08	35.95
	Realism	2.99	1.23	3	2.25	2.92	1.25	2.75	2.12	0.87	0.16	0.05	35.99
		μ	σ	M	IQR	μ	σ	M	IQR	p	z	r	-
NASA TLX	Mental demand	27.89	24.51	20	22.50	30.79	26.26	25	37.50	0.75	-0.32	0.05	
	Physical demand	15.00	17.87	5	5	16.58	19.15	5	25	0.63	0.48	0.08	
	Temporal demand	30.26	29.03	20	51.25	29.47	27.48	25	45	0.96	0.04	0.01	
	Performance	20.26	19.89	10	25	26.84	26.78	20	32.50	0.57	0.57	0.09	
	Effort	26.58	20.95	15	37.50	22.11	17.74	20	30	0.57	0.58	0.09	
	Frustration	12.89	16.69	5	5	17.11	19.74	5	18.75	0.45	0.76	0.12	
Behavioural intention		4.65	0.72	5	0.58	4.74	0.64	5	0.33	0.24	1.19	0.19	
QoE	Interface	3.90	0.88	4	1.75	3.84	0.83	4	1	1.00	0.00	0.00	
	Experience	4.05	0.85	4	1	3.84	0.83	4	1	0.16	-1.41	0.23	
Immersion	Immersion	3.79	1.93	4	3.88	3.76	1.93	4.5	3.5	0.93	0.09	0.01	
Presence	Presence	**4.74**	**1.91**	**6**	**2.75**	**4.26**	**1.88**	**4**	**3**	**0.02**	**2.26**	**0.37**	
	Spatial presence	4.23	1.33	4.5	2.62	4.19	1.32	5	2.44	0.86	-0.18	0.03	

participants (13) stated they liked and enjoyed the subtitle display interface: (*P13: "...So I found the interface easy to use, and I found the subtitles interface also easy to use and clear..."*), (*"P3: "...The interface, like the settings, was really intuitive and very straightforward. You see it, and you see exactly what's happening directly..."*). However, some participants (6) did not show the same acceptance: (*P12: "...So the subtitle interface, I did not like neither of them nor the 2D nor the 3D..."*). Some participants (15) were aware of the environment, expecting a low-poly version of the environment: (*P9: "...I felt immersive even in the in the prototype environment..."*), while others (4) were not aware and experienced a negative moment: (*P4: "...it's not consistent with the real world..."*).

In terms of subtitle placement, 8 participants preferred 3D (*P9: "...I prefer the dynamic subtitles because I can enjoy more the theatre so I can see the whole scene..."*), while 10 preferred 2D (*P8: "...because with the static, it was moving along with where I was looking, whereas in the dynamic, I had to look at the character speaking..."*). In the hypothetical scenario of a longer play, some participants (7) changed their opinion to the 2D subtitles (*P9: "...in this case, I would prefer static because it follows my head movement and after a long time, I will start to, to move my head and change my position ... I think the static position, it will be better. Especially if I use this on my home to lay down, I can see it..."*), while others (11) kept their opinion (*P2: "...still dynamic, I would still prefer it, especially if there were multiple actors and speakers..."*).

4 Testing subtitle placement in AR theatre

4.1 Design requirements

Results from the VR theatre testing showed no significant difference between the 2D and 3D subtitle placement, both considering quantitative results and qualitative insights. Thus, we sought out to evaluate how the subtitle placement would affect user experience in a real theatre environment, where subtitles and VFX would be experienced using AR glasses. Whereas the VR environment offered controlled, easily reproducible conditions, the live theatre offered different challenges to ensure a good experience: *(a) Subtitle synchronization*: The synchronization of the subtitles required live captioning of the actors' speech, to ensure correct timing of the subtitles appearing on-screen; *(b) Subtitle translation:* Subtitles should be made available in more languages other than English, to ensure wider accessibility; *(c) Subtitle placement*: The 3D subtitle placement required spatial analysis of the real environment, to be able to track the actors and follow their position accordingly. Similarly, the VFX placement also needed spatial analysis to be triggered in the correct positions.

To address the challenges, the following steps were taken:

- **Automated Speech Recognition (ASR).** While common techniques to achieve synchronization in captions rely on manual controls and human expertise, in this study, the captions were synced to the actor's speech by using speech transcription AI models and cross-referencing the resulting transcription against a set of pre-formatted captions. Post inference, additional quality assurance checks were implemented, for example by comparing

the inferred caption index to the previously delivered one, and by producing a visual notification for a human moderator to check the accuracy of the result if required.

- **Neural Machine Translation (NMT).** Translations of theatrical plays are typically undertaken by dedicated literary experts. Not only do they require high levels of experience and expertise, but they are also literary works of art in their own right. This is especially prevalent in Ancient Greek theatre plays. Given this, an AI approach for the translation of theatrical plays is only applicable in cases where no human translation of the play exists for the language in question. In the study, viewers were provided with clear demarcation of which languages were serviced by AI and which not. The AI ones were recommended if the viewers deemed that the benefit of having an AI translation in their own language was higher than having a human translation in a popular, foreign language such as English. In all cases though, the viewers were free to choose the language they preferred. To that end, the study created and provided AI translations of the play in four (4) languages (Dutch, Spanish, Italian and German) and provided the literary translation for the English and the (modern) Greek language. The AI generated translations were based of the English translation, rather than the Greek one, because of the improved performance of the translation model in languages with a larger available corpus. The fact that both the Greek and the English versions are translation of the Ancient Greek text (as opposed to translations of each other) further validates this approach. In this study, the translations were inferred in real-time by translating the pre-formatted, English caption using the previous two (2) captions as context for disambiguation and syntactical coherence.
- **Spatial Mapping.** 3D subtitles (and eventual VFX) need to be positioned in a specific location in the actual physical space and that position needs to be identical across all client applications, regardless of the viewer's position. To enable it, we selected to use a device (Magic Leap 2) and relative SDK that supports spatial mapping and object anchoring logic.

4.2 Experiment design and technical setup

The system architecture was set up with a server-client connection streaming messages in real time to AR glasses distributed to the audience. In the following section, we explain the main components that were added to the technical setup, namely, the ML models, and the adjustments made to the design and the equipment.

4.2.1 Machine learning (ML) Capabilities. The most critical component of the system is the machine learning model, running containerized using Docker technology. We perform caption provision in a cascaded fashion, where we start by transcribing the speech, then perform line matching of the transcripts to the captions. For the ASR component, we use Whisper [34] as a SOTA multilingual ASR model. Although Whisper supports up to 100 languages, its performance on low- to medium-resource languages such as Greek can be improved. To improve its performance on Greek language without affecting the other languages, we opt for adapter fine-tuning [36]. Adapter fine-tuning keeps the original model representations intact, but inserts small neural networks that are fine-tuned to specialize the model representations for the task, domain, or, in our case, the language. During inference, if the input language is Greek, we activate the adapters; if not, we ignore them. After fine-tuning the adapters, the Word Error Rate (WER) results went from 41.93% to 31.92% on Greek (lower is better). The caption matching accuracy was measured over the course of 12 theatrical performances by comparing the results of the AI inference with the performed script excerpt. Given clear signal and low noise levels, the system delivered an average of 86.46% accuracy over 2,474 samples.

Since ASR systems, especially in streaming settings, might not be reliable and faithful to the literary nature of theatre plays, we prepare subtitles of the play beforehand; then, during the play, we pick the most similar subtitle to the ASR transcription. To this end, we calculate the character n-gram F-score (chrF) [33] between the transcribed speech and all the lines in the subtitles in the input language. We select the entry with the highest score and return the line in the input language, the selected target language, and the score itself. The score can be used to disregard the provided translation if the score is below a predefined threshold. This ensures that we maintain high-quality subtitles while delivering low-latency transcription and translation.

The translations are prepared beforehand and are verified by a human expert to ensure accuracy. Given that most pre-trained translation models are trained on sentence-level translation, and thus, do not incorporate context, which is important in the literary scenario, we fine-tune NLLB-200 600M [42][1] for the task of Context-aware Machine Translation. We use the Huggingface Transformers library [44] for training. Specifically, we concatenate a single previous sentence (context) with the current sentence on the source language side, separated by the `[SEP]` token, and train the model to output only the current sentence on the target language side. This configuration prevents the compounding error caused by the imperfect translation of the source context as observed by Zhuocheng et al. [46]. To preserve the multi-lingual nature of the model, we selected six languages (English, Greek, German, Dutch, Italian, and Spanish), and trained the model for many-to-many translation between each of the six languages. The performance of the system was evaluated quantitatively by data logging from the server and the XR client applications during the AR Pilot. We used OpenSubtitles 2018 [24] dataset and randomly selected 100,000 examples for each language pair from the training subset. After fine-tuning, the model achieves higher scores for all language pairs on the majority of language pairs when context is utilized. Detailed results can be found in the supplemental materials. The overall latency included the sample duration, the transcription inference, translation inference (if requested by the user) and all the intermediate communication intervals. For a sample duration of 2s, transcription inference time displayed an average of 0.492s and a median of 0.491s, while overall latency (without translation) displayed an average of 2.498s and a median of 2.497s.

4.2.2 Application Design Interface. The user interface and the core application components were identical in the AR and VR versions (see Fig. 1). Their main difference is that the VR version included more visual elements to substitute for the lack of a physical scene and actors. The server application was running on a GPU-enabled laptop (Alienware X15 R2, Windows OS), positioned in the theatre

[1] https://huggingface.co/facebook/nllb-200-distilled-600M

hall. The NMT/ASR model was also running on the same laptop. The client application was running on two Magic Leap 2 devices. All devices were connected to the same WiFi network through a dedicated router (TP-Link TL- MR6400) set up in the same space. Both devices were connected with USB-C cables to laptops (one of which was the edge-server laptop). The laptops had the Magic Leap Hub application installed and the Device Stream enabled. The Device Stream module displayed the XR application contents to the laptop's monitor to allow observations by the research team and timely consultation or guidance to the users if necessary. Disinfection materials were provided with each XR device to the users along with instructions on disinfecting and wearing. For recording, a wireless microphone (RØDE Wireless ME) was used, whose transmitter was provided and attached to the actor, and whose receiver was connected via USB-C to the laptop running the server. The recording duration in the server was set to 2 seconds, and silence filtering was on with a threshold value of 0.05. The produced csv backlogs of the ASR transcription were stored locally in the edge-server laptop.

4.3 Experiment methodology and planning

4.3.1 Experiment methodology. Similarly to the previous experiment, a mixed-method methodology was chosen, comprised of quantitative questionnaires and semi-structured interviews. For the quantitative part, we tested for *Cognitive load* using the NASA TLX questionnaire [16], *Usability* using the SUS [8], *Behavioural Intention*, and *QoE*. We intentionally reduced the number of questionnaires to be answered by the users to reduce fatigue, excluding factors, such as *Presence* and *Immersion*, that have been designed specifically for VR environments. The *Behavioural Intention* questionnaire was expanded to include questions about machine translation. To account for cybersickness, we administered the SSQ before and after the experiment, following guidelines from the literature. The full questionnaires administered to the users are made available as supplemental material. For the qualitative part, the questions from the first part were modified to account for the complexity of the new setup, for example by inquiring about accuracy of captioning and of synchronization.

4.3.2 Participants. 12 participants (8 female, 4 male) from the general population participated in the AR experiment (3 participants in in the 18-24 age range, 3 participants in 25-34 age range, 5 participants in the 35-44 age range, 1 in 45-54 age range). On a scale from 1 (least) to 5 (most), the mean self-rated experience with immersive technologies was 2.2 (SD=1.1, median=2). The mean self-rated attendance in performances where string values were converted to numbers 1 (very rarely) to 5 (very often) was 3.7 (SD=1.1, median=4). Participants also self-assessed their familiarity with subtitles from 1 to 5, the mean was 2.6 (SD=1, median=3), showing medium use of subtitles. Preferentially, participants were non-native Greek speakers. Recruitment aimed to ensure diversity in age, gender, and XR familiarity to remove potential sampling bias. The majority of the participants choose English as the subtitle language; only two participants choose a different language. Post-hoc achieved power for the required effect size was 69%.

4.3.3 Procedure. Similarly with the VR experiment, a within-subject methodology was selected, in which all subjects experienced both the 2D and 3D variants of the subtitles, in randomized order. To effectively evaluate the impact of VFX, they were disabled in the first two steps, and a third step was added in which participants could freely choose between 2D and 3D subtitles, switching at will between the two within the experience. This allowed us to separately study the effect of VFX on the user experience, and simultaneously to record implicit preferences from the users regarding the subtitle placement, based on their choice.

At the start of the experiment, the researchers explained the purpose of the study, after which the participants had to sign a consent form to proceed with the test. Demographic data was also collected in this stage, and the SSQ was given to form a baseline of cybersickness. Then, the study proceeded as follows:

- **Step 0: tutorial and intro.** The participants (two per session) were welcomed in the theatre Alkmini in Athens, Greece. They were invited to sit on predefined spots in the second row and wear Magic Leap 2 glasses to enter the AR experience. They then experienced a tutorial on how to operate the glasses and the controllers.
- **Step 1: first theatre session.** Participants experienced the theatre play with either 2D or 3D subtitle placement and no VFX. The starting placement was randomized to avoid confounding factors. After the experience, they were asked to fill the NASA-TLX, and SUS. A break of 10 minutes was enforced to avoid fatigue.
- **Step 2: second theatre session.** Participants experienced the theatre play with either 3D or 2D subtitle placement, depending on which placement they had experienced in the previous step, and no VFX. After the experience, they were asked to fill the NASA-TLX and SUS. A break of 10 minutes was enforced to avoid fatigue.
- **Step 3: third theatre session.** Participants experienced the theatre play with 2D subtitles and VFX. At the beginning of the session, they were informed that they could change the subtitle placement at their will during the experience. After the experience, they were asked to fill the SSQ, NASA-TLX, SUS, Behavioural Intention, and QoE questionnaires. Following an optional short break (0-15 minutes), participants engaged in a semi-structured interview. The interview was audio-recorded. Upon conclusion, participants received debriefing information and were granted a voucher for their participation.

4.3.4 Data analysis. All questionnaires were grouped into the corresponding items based on the original papers. The quantitative data was tested for normality using the Shapiro-Wilk test (p values in supplemental material). Normally distributed data was checked for homogeneity of variance using Bartlett's test and for sphericity using Mauchly's test, and then analysed using repeated measures ANOVA. For non-normally distributed data, Friedman test was used. For qualitative results, the same procedure as in Sec. 3.2 was applied.

4.4 Results

4.4.1 Quantitative results. It can be observed in Table 2 that significantly lower SUS scores are obtained by the AR theatre application

with respect to the VR counterpart, for all cases. This indicates a drop to the third quartile, from the second quartile observed in VR [3]. This can be explained considering two factors: a) the AR application effectively merged reality and virtuality, thus posing more challenges in terms of usability - for example, due to errors of the spatial mapping of the application, or failures of the ML system; b) the audience of the AR experiment skewed more towards theatre-goers, and the experiment was conducted in a theatre, thus maybe raising expectations on the quality of the prototype. Shapiro-Wilk test indicated a normal distribution of the SUS scores, with no violation of homogeneity of variances and sphericity. Repeated measures ANOVA indicated no significant effect of runs on ratings, $F(2, 22) = 0.59, p = 0.56$, partial $\eta^2 = 0.05$.

Results indicate that the subtitle placement, either in 2D or 3D, led to low cognitive load. Shapiro-Wilk test indicated a normal distribution of factors Mental demand and Physical demand, with no violation of homogeneity of variances nor sphericity, whereas the other factors were non-normally distributed. Repeated measures ANOVA indicated no significant effect of runs on Mental demand, $F(2, 22) = 0.32, p = 0.73$, partial $\eta^2 = 0.03$, or on Physical demand, $F(2, 22) = 0.87, p = 0.43$, partial $\eta^2 = 0.07$. Similarly, Friedman test showed no significant effect of runs on Temporal demand, $\chi^2 = 0.19, p = 0.91$, Performance, $\chi^2 = 1.72, p = 0.42$, Effort, $\chi^2 = 1.17, p = 0.56$, and Frustration, $\chi^2 = 0.73, p = 0.70$.

In terms of behavioural intention, results indicated very favourable intention for the usage of machine translation ($\mu = 4.46, \sigma = 0.59$, max=5) and for AR theatre ($\mu = 4.42, \sigma = 0.87$, max=5). Similarly, the overall QoE was rated positively ($\mu = 4.25, \sigma = 0.45$, max=5). When prompted about which subtitle placement they preferred, users were divided: one person strongly preferred 3D to 2D subtitles, while the rest did not have a marked preference. This is also evidenced by the selected subtitle presentation in the third step: an equal number of users selected 2D and 3D subtitles when they had VFXs on (6 and 6), independently of the placement with which they started.

4.4.2 Qualitative results. Most of the participants expressed positive feelings regarding their participation in the AR theatre experience during their interviews (11). The general patterns observed in the interviews demonstrated that the idea of AR theatre was original (P11: *"I follow XR technology, mostly VR, so seeing it in theatre was something new and exciting"*) and immersive (P11: *"It was very immersive, it was visually impressive"*). Regarding the subtitle placement, an equal number of users expressed positive aspects of the 2D subtitle placement (7) (P05: *"it was less constrictive, I could move. I like to move my head a bit at the theater, and it gave me freedom"*), and of the 3D placement (7) (P08: *"I think the 3D, would help to understand who is speaking especially as I cannot hear [understand] the performance"*). Moreover, the majority of the participants expressed a positive reaction to the inclusion of visual and audio effects in the AR Theatre performance (10). Specifically, for the VFX, participants expressed clearly favourable opinions (7) and stated that they are interested in the potential for VFX in AR Theatre with more elaborate graphic design (3). A few participants characterized the inclusion of VFX as the most interesting part of the performance (P11: *"What stood out to me the most was the visual effects, especially the running horses. I loved that element!"*; P09: *"I really enjoyed the VFX. I was super excited when I saw the VFX coming out. And I think that they can really step up the entire experience and create a very, very immersive virtual environment around the actor and accentuate whatever they say and feel at that point"*). One of the participants specifically remarked on the inclusion of audio effects (P12: *"There were even sounds when something was happening! It was very interesting"*). They also highlighted the potential of AR Theatre specifically for ancient Greek plays, where the action always occurs off stage and is narrated through characters on stage (P10: *"For the first time we don't only imagine the narrative of Greek plays. With VFX, you could bring these scenes to life and the viewer can see it. I liked that a lot. You can have a powerful performance"*).

5 Discussion

5.1 2D and 3D subtitle placement

Our results showed no significant effect of subtitle placement on the user experience; in fact, we only saw a moderate effect of subtitle placement on the sense of presence in the VR case. In contrast with what was reported in previous studies [9, 22, 38], we saw no clear preference for either displaying mode, and its effect on workload or sickness was not significant. To limit the length of our experiment and reduce fatigue, we intentionally chose a short excerpt; longer plays might indicate a clearer preference for one modality versus the other. Moreover, as our environment was naturally focused on the stage, with fixed lighting conditions, the difference between the two modalities might have been less pronounced with respect to scenes with multiple points of focus, as is the case with omnidirectional video. Further studies are needed to better ascertain which subtitle placement might be optimal.

5.2 Technical limitations in AR theatre

The vast majority of the pain points identified by the participants of the AR experiment were related to technical aspects, issues, or failures. Some were related to the equipment used, like the AR glasses' restricted field of vision (5) (P06: *"I also felt like my field of vision was restricted, I had to move around to see everything, it was narrower than the human field of vision"*), or that the AR equipment was uncomfortable to use (4). However, the biggest limitations as experienced by the participants were related to the ML modules. Synchronization issues were experienced by all 12 participants, due to the latency incurred in the ASL module. Although the inference and communication values were fairly low on average, the required sample duration (2s) to achieve that performance is long enough to negatively affect the user experience, as reported by the users. For some of the participants the delay was significant enough to impact their enjoyment of the performance (5) (P07: *"I think I lost some things. It was distracting me a lot"*). Some participants also reported a change of pace in the delivery of the captions (4) (P04: *"It was like the subtitle was stuck for a few seconds and then the next subtitles run faster"*), or observed caption delivery issues, along the lines of displays of random/missing lines (4) (P05: *"Sometimes it displayed lines, but no one said them"*). Another issue lay in the automatic translation. The majority of the participants experienced the literary English translation; however, the two participants who experienced the NMT encountered severe issues both in terms of accuracy and caption delivery failure (error message displayed)

Table 2: Results of the questionnaires administered for the AR theatre experiment, for 2D, 3D, and VFX run.

		2D				3D				VFX			
		μ	σ	M	IQR	μ	σ	M	IQR	μ	σ	M	IQR
SUS		69.79	19.38	72.5	26.25	67.71	18.72	68.75	28.75	65.63	22.64	63.75	33.75
NASA-TLX	Mental demand	31.67	21.25	22.5	27.5	33.75	19.90	30	20	36.67	25.08	37.5	35
	Physical demand	22.92	14.99	20	27.5	18.75	17.60	12.5	20	25.00	18.95	25	27.5
	Temporal demand	24.58	20.39	15	37.5	25.83	19.87	20	30	24.58	16.16	25	27.5
	Performance	25.00	18.95	20	37.5	32.92	25.80	32.5	42.5	27.50	27.51	12.5	45
	Effort	19.17	14.28	17.5	15	25.42	18.88	27.5	27.5	28.33	24.89	25	40
	Frustration	22.08	18.64	15	30	25.83	22.95	17.5	42.5	27.50	24.45	15	42.5

and both also experienced all three of the general caption issues described in the previous paragraph (P06: *"I was not understanding what was happening in the monologue. Sometimes there were 500 errors instead of translation. Sometimes they went really fast too, line after line"*). Both participants switched language to English after the completion of the first run of their session. The limitations are due to the strain of running accurate ML models while keeping latency low for real-time consumption, and point to a large issue that needs to be overcome in order to enable deployment of AR subtitling at scale, where latency issues might be exacerbated [35]. NMT models can be an important tool to democratize access to plays in a foreign language, especially one in which no official translation is present. However, such languages would incur into the problem of not having large corpora on which to be trained [14], which might negatively impact the translation quality and, in turn, further alienate the audience instead of bringing it closer. Compromising accuracy to reduce latency might not be possible in these scenarios.

5.3 Accessibility potential

Subtitles allow for broader accessibility of video content to diverse populations of viewers, such as non-native language speakers, the hard of hearing, and people with learning disabilities [25]. Personalized AR subtitles have the potential for enhanced accessibility, as they can offer multiple translations simultaneously while remaining unobtrusive. Such potential was recognized by our subject pool: participants (8) pointed out the potential of AR captions in increasing accessibility to theatre performances, in terms of watching foreign performances (P10: *"It makes it a big deal. The ability to be able to follow the performance or not"*). Most comments on the usefulness of the captions naturally came from non - native/ non- fluent Greek speakers, however the potential of the captions was recognized by Greek speakers. P02 noted that the captions themselves may serve people learning the language of the performance (P02: *"I can go and see a Greek performance while not understanding fully Greek. Then first, I'm learning and second, I'm understanding what I'm seeing"*). Two participants remarked on the advantages of AR subtitles over the standard practice of delivering captions through screens during a performance, either in terms of comfort/immersion or accessibility to audiences with learning disabilities (P02: *"I am very dyslexic also, so it was fine for me to follow"*). Moreover, some participants highlighted the potential for the spread of AR captions technology (P12: *"This has the potential to be a game changer for the theatre experience. The subtitles are the most promising feature, in the future I could see them get adopted by many theatres"*).

5.4 Challenges and possibilities for AR Theatre

While AR theatre primarily enhances the audience experience, it also significantly impacts actors, directors, and other theatre practitioners who are neither the direct users of AR technology nor its core developers , positioned between traditional theatrical practice and emerging digital tools. Our discussion with our theatre partners uncovered several challenges and opportunities for their craft:

- *For Actors Performing in an Augmented Space.* Actors must adapt to a performance space where audience perception is no longer uniform. Some viewers will experience added VFX, subtitles, or translated audio, while others will engage with the play in its raw, traditional form. This creates dual-performance challenges: (a) Maintaining consistent emotional and physical expression despite varying audience interpretations; (b) Reacting naturally to unseen digital elements (e.g., a VFX-enhanced moment or AR-generated cues); (c) Spatial awareness adjustments, ensuring movements align with both the live audience and the AR-enhanced experience. Additionally, blocking and stage movement may be affected if AR overlays require actors to stay within specific "AR-friendly" zones to avoid visual occlusion.
- *For Directors: Balancing Classic Theatrical Expression and Innovation.* Directors could rethink narrative structure and staging to accommodate AR's presence while preserving theatrical authenticity. Key considerations include (a) Choreographing AR effects so they support rather than distract from the performance; (b) deciding the level of audience agency—how much control should viewers have over subtitles, translations, or visual elements? (c) Synchronizing live performance with digital effects, ensuring timing between actors and AR-generated content is seamless; (d) Rehearsal adaptations, requiring actors and stage managers to incorporate AR elements into practice sessions, even if they are not visible during rehearsals. This new medium forces directors to think beyond a single visual plane, embracing multi-perspective storytelling.
- *For Set Designers and Technical Teams Integrating AR into Physical Space.* Unlike traditional theatre, where set pieces and lighting define the visual world, AR introduces a new digital layer that must blend seamlessly with the physical environment. This affects (a) Stage design, ensuring AR visuals don't clash with physical set pieces; (b) Lighting design, as AR elements need proper visibility without washing out the projections; (c) Costume considerations—how actors' attire interacts with AR overlays, especially if colour tracking or marker-based interactions are used. AR theatre is not merely a technological upgrade—it reframes the

creative process. The challenge is to retain the soul of live performance while embracing the expanded possibilities AR offers. Theatres have historically adapted to electric lighting, microphones, and projections — AR may simply be the next step in this evolution. However, its adoption must be artist-driven, not technology-imposed, ensuring that storytelling remains at the heart of the experience.

6 Conclusion

We presented two studies aimed at uncovering how subtitle placement affects user experience in a VR/AR Theatre setting. In particular, we first tested two modes of subtitle placement in a virtual play, which allowed us to evaluate the user experience in a controlled environment, and then we tested the same conditions in an AR setting, where we integrated automated captioning and machine translation. Our study provided valuable insights into how participants react to VR and AR Theatre experiences.

Acknowledgments

The authors want to thank Abdallah El Ali for encouraging them in the submission process. This work was supported through the Horizon Europe research and innovation programme, under grant agreement No 101070521 (VOXReality).

References

[1] A. Brown, J. Turner, J. Patterson, A. Schmitz, M. Armstrong, and M. Glancy. 2018. Exploring Subtitle Behaviour for 360° Video. BRITISH BROADCASTIND CORPORATION, 40. https://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP330.pdf

[2] António Baía Reis and Mark Ashmore. 2022. From video streaming to virtual reality worlds: an academic, reflective, and creative study on live theatre and performance in the metaverse. *International Journal of Performance Arts and Digital Media* 18, 1 (2022), 7–28.

[3] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[4] Sharon Black. 2022. Subtitles as a Tool to Boost Language Learning and Intercultural Awareness? : Children's Views and Experiences of Watching Films and Television Programmes in Other Languages with Interlingual Subtitles. *Journal of Audiovisual Translation* 5, 1 (April 2022), 73–93. https://doi.org/10.47476/jat.v5i1.2022.157 Number: 1.

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* (Jan. 2006). https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa Publisher: Taylor & Francis Group.

[6] M. Brescia-Zapata, K. Krejtz, A. Duchowski, and C. Hughes. 2022. VR 360° subtitles: designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation* 6, 2 (Dec. 2022). https://salford-repository.worktribe.com/output/1324830/vr-360o-subtitles-designing-a-test-suite-with-eye-tracking-technology

[7] Marta Brescia-Zapata, Krzysztof Krejtz, Andrew T. Duchowski, Christopher J. Hughes, and Pilar Orero. 2023. Eye-tracked Evaluation of Subtitles in Immersive VR 360° Video. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 769–770. https://doi.org/10.1109/VRW58643.2023.00227

[8] john Brooke. 1996. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press. Num Pages: 6.

[9] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. 2015. Dynamic Subtitles: The User Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video (TVX '15)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2745197.2745204

[10] Michael Crabb, Rhianne Jones, and Mike Armstrong. 2015. The Development of a Framework for Understanding the UX of Subtitles. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 347–348. https://doi.org/10.1145/2700648.2811372

[11] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[12] Deborah I. Fels. 2002. Accessible Digital Media. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Joachim Klaus, and Wolfgang Zagler (Eds.). Springer, Berlin, Heidelberg, 282–283. https://doi.org/10.1007/3-540-45491-8_59

[13] Victor Rogério Sousa Ferreira, Lisle Faray de Paiva, Anselmo Cardoso de Paiva, Régis Costa de Oliveira, Mónica Sofia Santos Mendes, and Ivana Marcia Oliveira Maia. 2020. Authoring and Visualization Tool for Augmented Scenic Performances Prototyping and Experience. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*. IEEE, 413–419.

[14] Steven Fincke, Shantanu Agarwal, Scott Miller, and Elizabeth Boschee. 2022. Language model priming for cross-lingual event extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 10627–10635.

[15] Cyrielle Garson. 2024. New Community Design to the Rescue: The Promises and Pitfalls of Post-Pandemic VR Theatre in North America. *Journal of Contemporary Drama in English* 12, 2 (2024), 249–269.

[16] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (Oct. 2006), 904–908. https://doi.org/10.1177/154193120605000909 Publisher: SAGE Publications Inc.

[17] Melinda Hestiana and Anita Anita. 2022. THE ROLE OF MOVIE SUBTITLES TO IMPROVE STUDENTS' VOCABULARY. *Journal of English Language Teaching and Learning* 3, 1 (July 2022), 46–53. https://doi.org/10.33365/jeltl.v3i1.1715 Number: 1.

[18] Sarah Hudson, Sheila Matson-Barkat, Nico Pallamin, and Guillaume Jegou. 2019. With or without you? Interaction and immersion in a virtual reality experience. *Journal of Business Research* 100 (July 2019), 459–468. https://doi.org/10.1016/j.jbusres.2018.10.062

[19] 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 (DIS '18 Companion)*. Association for Computing Machinery, New York, NY, USA, 7–11. https://doi.org/10.1145/3197391.3205404

[20] Alkiviadis Katsalis, Konstantinos Christantonis, Charalampos Tsioustas, Pantelis I Kaplanoglou, Maximos Kaliakatsos-Papakostas, Athanasios Katsamanis, Konstantinos Diamantaras, Vassilis Katsouros, Evita Fotinea, Depy Panga, et al. 2022. NLP-Theatre: Employing Speech Recognition Technologies for Improving Accessibility and Augmenting the Theatrical Experience. In *Proceedings of SAI Intelligent Systems Conference*. Springer, 507–526.

[21] Iryna Kuksa and Mark Childs. 2014. 4 - Theatre in the virtual day and age. In *Making Sense of Space*, Iryna Kuksa and Mark Childs (Eds.). Chandos Publishing, 51–65. https://doi.org/10.1533/9781780634067.2.51

[22] Kuno Kurzhals, Emine Cetinkaya, Yongtao Hu, Wenping Wang, and Daniel Weiskopf. 2017. Close to the Action: Eye-Tracking Evaluation of Speaker-Following Subtitles. 6559–6568. https://doi.org/10.1145/3025453.3025772

[23] Sueyoon Lee, Moonisa Ahsan, Irene Viola, and Pablo Cesar. 2025. User centric Requirements for Enhancing XR Use Cases with Machine Learning Capabilities. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 753–757.

[24] Pierre Lison, Jörg Tiedemann, and Milen Kouylekov. 2018. OpenSubtitles2018: Statistical Rescoring of Sentence Alignments in Large, Noisy Parallel Corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga (Eds.). European Language Resources Association (ELRA), Miyazaki, Japan. https://aclanthology.org/L18-1275

[25] Shadi R. Masadeh and Saba A. Soub. 2020. Educational Video Classification Using Fuzzy Logic Classifier Based on Arabic Closed Captions. In *Intelligent Computing Paradigm and Cutting-edge Technologies (Learning and Analytics in Intelligent Systems)*, Lakhmi C. Jain, Sheng-Lung Peng, Basim Alhadidi, and Souvik Pal (Eds.). Springer International Publishing, Cham, 133–138. https://doi.org/10.1007/978-3-030-38501-9_13

[26] Miguel Melo, Guilherme Gonçalves, josé Vasconcelos-Raposo, and Maximino Bessa. 2023. How Much Presence is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score. *IEEE Access* 11 (2023), 24675–24685. https://doi.org/10.1109/ACCESS.2023.3254892 Conference Name: IEEE Access.

[27] Estella Oncins, Rocío Bernabé, Mario Montagud, and Verónica Arnáiz Uzquiza. 2020. Accessible scenic arts and virtual reality: a pilot study with aged people about user preferences when reading subtitles in immersive environments. (2020).

[28] De Xing Ong, Kai Xiang Chia, Yi Yi Huang, Jasper Teck Siong Teo, Jezamine Tan, Melissa Lim, Dongyu Qiu, Xinxing Xia, and Frank Yunqing Guan. 2021. Smart captions: a novel solution for closed captioning in theatre settings with AR glasses. In *2021 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, 1–5.

[29] ChangHoon Park, Heedong Ko, Ig-Jae Kim, Sang Chul Ahn, Yong-Moo Kwon, and Hyoung-Gon Kim. 2002. The making of Kyongju VR theatre. In *Proceedings IEEE Virtual Reality 2002*. IEEE, 269–270.

[30] Pranav Pidathala, Dawson Franz, James Waller, Raja Kushalnagar, and Christian Vogler. 2022. Live Captions in Virtual Reality (VR). https://doi.org/10.48550/arXiv.2210.15072 arXiv:2210.15072 [cs].

[31] Krzysztof Pietroszek, Manuel Rebol, and Becky Lake. 2022. Dill Pickle: Interactive Theatre Play in Virtual Reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (VRST '22)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3562939.3565678

[32] Shane Pike. 2020. Virtually relevant: AR/VR and the theatre. *Fusion Journal* 17 (2020), 120–128.

[33] Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, Ondřej Bojar, Rajan Chatterjee, Christian Federmann, Barry Haddow, Chris Hokamp, Matthias Huck, Varvara Logacheva, and Pavel Pecina (Eds.). Association for Computational Linguistics, Lisbon, Portugal, 392–395. https://doi.org/10.18653/v1/W15-3049

[34] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356 [eess.AS] https://arxiv.org/abs/2212.04356

[35] Mohaimenul Azam Khan Raiaan, Md Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad, Sadman Sakib, Most Marufatul Jannat Mim, Jubaer Ahmad, Mohammed Eunus Ali, and Sami Azam. 2024. A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE access* 12 (2024), 26839–26874.

[36] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. 2017. Learning multiple visual domains with residual adapters. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/e7b24b112a44fdd9ee93bdf998c6ca0e-Paper.pdf

[37] Estelle Rivier-Arnaud. 2020. Doran's and Taymor's Tempests: Digitalizing the Storm, a Dialogue between Theatre and Cinema. *Représentations dans le monde anglophone* 21 (2020).

[38] Sylvia Rothe, Kim Tran, and Heinrich Hußmann. 2018. Dynamic Subtitles in Cinematic Virtual Reality. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '18)*. Association for Computing Machinery, New York, NY, USA, 209–214. https://doi.org/10.1145/3210825.3213556

[39] Augusto Rupérez Micola, Ainoa Aparicio Fenoll, Albert Banal-Estañol, and Arturo Bris. 2019. TV or not TV? The impact of subtitling on English skills. *Journal of Economic Behavior & Organization* 158 (Feb. 2019), 487–499. https://doi.org/10.1016/j.jebo.2018.12.019

[40] Rufat Rzayev, Sabrina Hartl, Vera Wittmann, Valentin Schwind, and Niels Henze. 2020. Effects of position of real-time translation on AR glasses. In *Proceedings of Mensch und Computer 2020 (MuC '20)*. Association for Computing Machinery, New York, NY, USA, 251–257. https://doi.org/10.1145/3404983.3405523

[41] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. 2018. Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3173574.3173619

[42] NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv:2207.04672 [cs.CL] https://arxiv.org/abs/2207.04672

[43] Abigail LM Webb, Paul B Hibbard, Jessica Dawson, Loes C Van Dam, Jordi M Asher, and Leo J Kellgren-Parker. 2024. Immersive-360° theater: User experience in the virtual auditorium and platform efficacy for current and underserved audiences. *Psychology of Aesthetics, Creativity, and the Arts* (2024).

[44] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Qun Liu and David Schlangen (Eds.). Association for Computational Linguistics, Online, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[45] Aleksandra Zheleva, Anne Roos Smink, Paul Hendriks Vettehen, and Paul Ketelaar. 2021. Modifying the Technology Acceptance Model to Investigate Behavioural Intention to Use Augmented Reality. In *Augmented Reality and Virtual Reality*, M. Claudia tom Dieck, Timothy H. Jung, and Sandra M. C. Loureiro (Eds.). Springer International Publishing, Cham, 125–137. https://doi.org/10.1007/978-3-030-68086-2_10

[46] Zhang Zhuocheng, Shuhao Gu, Min Zhang, and Yang Feng. 2023. Scaling Law for Document Neural Machine Translation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 8290–8303. https://doi.org/10.18653/v1/2023.findings-emnlp.556