https://doi.org/10.58256/64dz1c72





RESEARCH ARTICLE **Section:** Language and Linguistics Assessing Al-driven dubbing websites: Reactions of Arabic native speakers to Al-dubbed English videos in Arabic

Maysa Ahmad, Ahmad S Haider* 6 & Hadeel Saed

Applied Science Private University, Jordan *Correspondence: a haidar@asu.edu.jo

ABSTRACT

In the era of rapid Artificial Intelligence (AI) advancements, emerging tools have become essential components of our daily routine. AIdubbing aims to speed up localisation by replacing the original soundtrack with AI-generated sounds. These developments raise the question of whether human dubbers could be replaced in the future. This study quantitatively examines viewers' reactions to the AI Arabicdubbed versions of the "Pride and Prejudice" movie using AI dubbing tools from two websites: ElevenLabs and Dübverse. The participants are asked to watch the original video along with two dubbed videos generated by the two websites. For data collection purposes, a three-point Likert scale questionnaire consisting of 19 items and five constructs —linguistic issues, technical issues, synchronisation, translation quality, and attitudes and future recommendations -was designed to elicit the reactions of 150 native Arabic speakers. The analysis shows that linguistic issues, technical issues, synchronisation, and translation quality significantly affected the participants' attitudes and future recommendations regarding the use of AI-dubbing services. 80% of the respondents acknowledged that AI-dubbing is beneficial for making content accessible to a wider audience. The study found that ElevenLabs outperforms Dübverse, especially in areas such as voice cloning, maintaining both kinesic and isochrony, handling colloquial language, managing multiple speakers, and achieving overall better performance in translation. The findings showed that both websites lack lip-synchrony and require enhancements in other areas. This study is beneficial for content creators seeking to expand their reach globally. The study recommends conducting further research on AI-dubbing across different genres and languages.

KEYWORDS: Arabic, Artificial Intelligence, Audio-visual Translation (AVT), Dubbing, English, Series

Research Journal in Advanced Humanities

Volume 6, Issue 1, 2025 ISSN: 2708-5945 (Print) ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: 22 January 2025 Accepted: 11 March 2025 Published: 10 April 2025

HOW TO CITE

Ahmad, M. ., Haider, A. S., & Saed, H. . (2025). Assessing AIdriven dubbing websites: Reactions of Arabic native speakers to AIdubbed English videos in Arabic. Research Journal in Advanced *Humanities*, 6(1). https://doi. org/10.58256/64dz1c72



Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited © 2025 The Author(s). This is an open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Study Background

Advances in technology are demonstrating the potential achievements and the possibilities that can be unlocked through AI tools. In recent years, there has been a demand for localised content from viewers and content creators, which has driven significant growth in the global dubbing and localisation market. As AI-driven dubbing websites become a cost-effective and efficient alternative to traditional methods, a critical question arises: Can AI genuinely replace the human touch required for effective dubbing? Although deep learning algorithms enable AI to mimic the flow and rhythm of human speech, concerns persist regarding its ability to capture and convey the complex emotions that characterize human speech. Additionally, the complex grammatical structures, linguistic nuances, and language varieties inherent in languages like English and Arabic pose significant challenges for AI-driven dubbing machines and raise concerns about the fidelity and naturalness of dubbed content.

Different studies have investigated traditional dubbing in audiovisual translation (AVT) across various languages (Alrousan & Haider, 2022; Hayes, 2021; Iitola, 2017; Yahiaoui, 2016). However, few studies investigated automatic dubbing (Baños, 2023; Brannon, Virkar, & Thompson, 2023; Federico, Virkar, Enyedi, & Barra-Chicote, 2020). To the best of the researchers' knowledge, no studies have been conducted related to evaluating the performance of AI-driven dubbing tools using Elevenlabs and Dübverse websites, particularly in the context of speech-to-speech (STS) translation from English into Arabic. Accordingly, this study fills this gap by evaluating the performance of AI in dubbing videos across different genres using the two websites. In addition, this work investigates viewers' reactions towards AI-dubbed material and how they perceive the naturalness and authenticity of such AI-dubbed content.

This study addresses the following three questions:

- To what extent do AI-driven dubbing websites (ElevenLabs and Dübverse) achieve fluency, accuracy, and stylistic nuance when translating English videos into Arabic?
- How do ElevenLabs and Dübverse technically compare when dubbing English videos into Arabic in terms of synchronisation, natural-sounding intonation, voice selection, voice performance, and appropriate sound quality?

2. Review of Related Literature

2.1. Al-Dubbing

Chaume (2020, p. 104) defines dubbing as "a linguistic, cultural, technical and creative team effort that consists of translating, adapting and lip-syncing the script of an audiovisual text." Bigioi and Corcoran (2023) discuss that traditional dubbing is a costly post-production process consisting of three primary steps. First, translation is carried out by language experts who are proficient in the source and the target languages. Second, voice acting necessitates identifying and hiring suitable voice actors who can effectively portray the emotions of the original recording and match the on-screen character. Lastly, audiovisual mixing combines the newly recorded audio voice with the original video recording. This step involves manual editing to synchronise the new audio with the original video properly. This process is costly and time-consuming, but it would greatly benefit from automation. Thanks to recent advancements in deep learning (DL), opportunities have opened up to automate each of the traditional dubbing steps.

Automatic Dubbing (AD) is defined as "the task of automatically replacing the speech in a video document with the speech in a different language while preserving as much as possible the user experience of the original video" (Lakew et al., 2021, p. 1). AD and AI dubbing are two terms that refer to the application of machine learning (ML) techniques and advanced algorithms to achieve the dubbing process. However, "Automatic dubbing" might be a broader term that includes any automated process of replacing audio in a video, not necessarily using AI specifically. "AI-dubbing" explicitly emphasises the use of AI technology in the process, such as speech synthesis, natural language processing (NLP), DL and MT, to perform tasks such as translation, speech synthesis and voice recognition.

An essential requirement of dubbing is speech synchronisation (see Table 1), which should happen at the

lip movement level (lip synchrony), at the utterance level (isochrony), and at the body movement level (kinesic synchrony) (Chaume, 2004). Most of the research on AD addressed the isochrony, aiming to generate utterances that align with the phrase-pause arrangement of the original audio (Federico, Virkar, et al., 2020; Lakew et al., 2021; Öktem, Farrús, & Bonafonte, 2019).

	Synchrony type	Definition
1	1 Kinesic synchrony	Following the gestures, the meaning of gestures has to be in consonance with the
1		translation; translation is subordinate to the intention of gestures.
2	Isaahaany	Following the timing of each character's speech, every sentence, phrase, and pause
2	isochrony	has to coincide with the timing needed by an actor to utter them.
2	Lingung	it is only an issue in close-up shots, where the speaker's face and lip movements are
3	пр-зупс	fully visible, and only in labials and semi-labials, where the mouth has to be closed

 Table 1: Synchrony types (Chaume, 2004)

According to Wahlster (2013), AD is considered an extension of speech-to-speech translation (STS) translation and text-to-speech (TTS) translation. STS translation is translating a speech in one language to a speech in another language without relying on an intermediate text generation (Lee et al., 2021). STS translation is a combination of three technologies: ASR, used to transcribe speech to text (STT) in the source language (SL); MT, which translates the transcribed text to the target language (TL); and TTS, which converts the MT output into speech (see Figure 1).

Figure 1: Speech translation pipeline to perform AD enhancements (Lakew et al., 2021).



AI-powered systems analyse spoken input, translate it into the desired language, and generate synthesised speech that closely mimics natural human speech. These systems facilitate multilingual communication across various settings, including business meetings, international conferences, and travel scenarios. However, AD differs from STS translation in that the primary objective of STS translation is to generate an output reflecting the original speech's linguistic content. AD "aims to replace all speech contained in a video document with the speech in a different language so that the result sounds and looks as natural as the original" (Federico, Virkar, et al., 2020, p. 257). Besides this, dubbing should also match the original tone, emotion, duration, background noise, and prosody.

TTS translation is defined as "a process in which input text is first analysed, then processed and understood, and then the text is converted to digital audio and then spoken" (Trivedi, Pant, Shah, Sonik, & Agrawal, 2018, p. 39). Moreover, they discussed the stages of the TTS translation conversion. Figure 2 illustrates all the stages involved in TTS conversion; however, the primary stages of TTS systems are text processing and speech synthesis.

Figure 2: Text-to-speech system flow (Trivedi et al., 2018)



In the text-processing stage, the input text is analysed and transcribed into linguistic and phonetic representations. The speech synthesis process encompasses several categories, including articulator synthesis, which utilizes both mechanical and acoustic models to generate speech. Formant synthesis shows individual speech segments. Concatenative synthesis synthesises sound by combining short samples of sound known as units.

2.2. Empirical Studies

This section reviews some empirical studies on traditional dubbing and automated dubbing in different languages. Alrousan and Haider (2022) examined the strategies used by translators in dubbing English television advertisements into Arabic. Six advertisements from various domains were selected from YouTube, and their English and Arabic versions were compared in terms of the contents, brand names, and catchphrases. The results showed that the translators employed various strategies to accurately and effectively convey the original message, including cultural adaptation, loan translation, substitution, transliteration, explicitation, addition, omission, and paraphrasing. It is also noted that there were instances where translators opted for literal translation to maintain the foreignness spirit and fidelity to the original message.

Additionally, Abulawi, Al-Salman, and Haider (2022) investigated the use of Modern Standard Arabic (MSA) and the Egyptian Vernacular (EV) in dubbing animated movies in the Arab world, taking A Bug's Life as a case study. This study employs a mixed-methods approach, incorporating both qualitative and quantitative methods. The researchers qualitatively investigated the data by compiling a parallel corpus using a corpusbased software package to analyse the data and spot the differences between the two varieties. Quantitatively, a questionnaire was constructed to elicit the participant's reactions to the given scenes. The qualitative analysis revealed that the EV dubbed version surpasses its MSA counterpart in terms of creativity, coherence, conveying emotion, maintaining meaning, and respecting the target audiences' cultures. The quantitative analysis indicated that the participants strongly preferred the EV dubbed version over MSA. They found it more entertaining. Furthermore, the technical aspects, including voice quality and synchrony, are better reflected in the EV version than in the MSA one.

Federico, Enyedi, et al. (2020) investigated the enhancement of a speech-to-speech (STS) translation pipeline for automatic dubbing (AD), integrating prosodic alignment, neural machine translation (NMT), neural text-to-speech (TTS) with duration tuning, and audio rendering. The study employs a mixed-methods approach, combining both qualitative and quantitative methods. Qualitatively, the Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA) methodology was adopted to assess the naturalness of dubbed TED Talk excerpts translated from English into Italian. Quantitatively, 24 video clips extracted from 6 TED Talks (3 female and 3 male speakers, with 5 clips per talk) were rated by Italian and non-Italian listeners to assess the naturalness of the language using a 0-100 scale. The participants were instructed to disregard the content

and focus only on the naturalness of the output. The results showed that the prosodic alignment had a negative impact on fluency and prosody, especially for Italian listeners, whereas audio rendering notably enhanced the experience for non-Italian listeners. Italian participants showed a preference for the baseline system, which is the system without any additional enhancements or modifications. In contrast, non-Italian participants leaned towards the enhanced versions that had undergone modifications or improvements.

Baños (2023) explored the challenges of using automatic dubbing to translate educational YouTube videos posted in the Spanish version of the YouTube Channel Amoeba Sisters, which have been dubbed using the Aloud tool with minimal human intervention. The study focused on the involvement of Machine Translation (MT), Automatic Speech Recognition (ASR), and Text-to-Speech (TTS) technologies to automatically replace the original audio tracks with synthetic speech in another language. The results showed that improvements are still needed in terms of the naturalness of synthetic voices, accuracy, fluency, and delivery pace.

Chronopoulou et al. (2023) proposed training a model that optimised both the translation quality and the speech duration of the generated translations simultaneously to enhance isochrony in automatic dubbing. This study is quantitative. A standard encoder-decoder transformer model was trained to predict phonemes and their duration in the target language. The researchers evaluated the effectiveness of the proposed model by conducting experiments using the Fairseq library. The results showed a significant enhancement in speech overlap compared to baseline models (StdMT, IsoMT, and Txt2Phn), with a 55% relative increase. Additionally, the researchers released a new dubbing test set for use by researchers in the dubbing community.

Although limited studies have examined AI dubbing, to the best of the researchers' knowledge, no studies have investigated dubbing English videos into Arabic using diverse AI tools, particularly in terms of fluency, accuracy, and style. Therefore, this study fills this gap by evaluating the performance of ElevenLabs and Dübverse websites in dubbing AV content from English to Arabic.

3. Methodology

3.1. Investigated Dubbing Websites

Dubbing has traditionally been associated with high costs and a lengthy process due to the extensive resources required for studio recording sessions, including equipment, devices, technicians, and voice and language talents. Moreover, it takes time to dub and edit the content. However, advancements in technology have revolutionised and given rise to new technology known as AI-dubbing technology, which offers affordable solutions that make the process more flexible and cost-effective. ElevenLabs and Dübverse websites, which are paid services, represent a diverse range of AI-dubbing tools, each offering unique features and capabilities for translation tasks. According to MurfResources' (2023) ranking of the top eight AI-dubbing websites, ElevenLabs ranked first, and Dübverse ranked third. Having a high ranking and unique features, including Arabic and English language options, influenced the decision to investigate them. By conducting a comparative analysis of these websites, this study assesses the performance, accuracy, and suitability of various AI-dubbing tools for translating diverse audiovisual materials from English into Arabic.

3.2. Questionnaire Validity

For the purpose of this study, a questionnaire was constructed and distributed to 150 native Arabic speakers whose major field of study is English to gather their responses regarding the Arabic AI-dubbing of English videos. The questionnaire consisted of two sections. The first section sought to obtain demographic information about the participants. The second section examined their reactions to the Arabic AI-dubbed videos using ElevenLabs and Dübverve, and sought future recommendations for enhancing the quality of AI-dubbing practices. The analysis adopted the three-point Likert scale of: (1) Agree; (2) Uncertain ; (3) Disagree. In this part, the participants' responses were statistically analysed.

After consulting several similar questionnaires, the researchers developed the research instrument for this study. The questionnaire was reviewed by a jury of three translation experts, who provided comments and feedback to ensure its validity. Their feedback and suggestions were incorporated accordingly before the questionnaire was distributed. As previously noted, the questionnaire was structured into two main sections. The first section collected demographic data, including participants' gender, age, education, and English proficiency. The second section of the questionnaire consisted of 19 items distributed across five constructs: Linguistic

Issues, Technical Issues, Synchronisation, Translation Quality, and Attitudes and Future Recommendations.

3.3. Questionnaire Reliability

To ensure that the 19 statements were clear, consistent, and closely related to the main objective of this study, the researchers conducted a pilot study with 25 participants to assess the questionnaire's reliability. Additionally, reliability was further examined using Cronbach Alpha results and correlation coefficients. Moreover, Table 2 presents the results of the questionnaire's 19 statements and their relationship to the five subscales.

Construct	No. of Items	Cronbach's Alpha
Linguistic Issues	3	.773
Technical Issues	6	.903
Synchronisation	3	.819
Translation Quality	3	.913
Attitudes and Future Recommendations	4	.782
All Variables	19	.926

Table 2: Reliability Analysis through Cronbach Alpha Results (Cronbach, 1951)

Table 2 indicates a Cronbach's alpha coefficient of 0.912 for all variables, indicating a high level of reliability for the questionnaire and strong internal consistency. Concerning the results of each construct, including Linguistic Issues, Technical Issues, Synchronisation, Translation Quality, and Attitudes and Future Recommendations, Cronbach's alpha values were equal to 0.773, 0.903, 0.819, 0.913, and 0.782, respectively. According to Nunnally (1978), a reliability coefficient of 0.70 or higher is considered "acceptable" in social science research.

3.4. Normality Test

The researchers conducted skewness and kurtosis tests to ensure that the data were normally distributed, as shown in Table 3.

Table 3: Normality indicator

Construct	Skewness	Kurtosis
Linguistic Issues	0.973	0.133
Technical Issues	1.382	0.705
Synchronisation	1.477	1.129
Translation Quality	1.008	-0.512
Attitudes and Future Recommendations	-1.145	2.179

The data is deemed to follow a normal distribution as the skewness and kurtosis values in Table (3) fall within the acceptable range of (-3) to (+3), indicating that the data is normally distributed (George & Mallery, 2003).

3.5. Measurement Model

The measurement model, consisting of five constructs assessed by 19 measurement items, was evaluated using confirmatory factor analysis (CFA) in AMOS. The analysis aimed to validate the factor structure of the observed variables, including the factor loadings. Composite reliability (CR) and convergent validity were also assessed. The results are presented in Table 4. Discriminant validity was assessed using HTMT Analysis, and the results are provided in Table (5).

 Table 4: Confirmatory factor analysis results

Latent Variable	Indicator	FL	FLS	AVE (> 0.50)	CR (> 0.70)	
	LI1	0.734	0.539			
Linguistic issues	LI2	0.778	0.605	0.522	0.765	
	LI3	0.649	0.421			
	TI1	0.777	0.604			
	TI2	0.84	0.706			
Technical Issues	TI3	0.734	0.539	0.(22	0.012	
	TI4	0.841	0.707	0.633	0.912	
	TI5	0.829	0.687			
	TI6	0.746	0.557			
	SY1	0.756	0.572			
Synchronisation	SY2	0.719	0.517	0.587	0.810	
	SY3	0.82	0.672			
	TQ1	0.796	0.634			
Translation Quality	TQ2	0.838	0.702	0.668	0.858	
	TQ3	0.817	0.667			
	AFR1	0.543	0.295			
Attitudes and Entrino Decommon deticing	AFR2	0.768	0.590	0.527	0.014	
Attitudes and Future Recommendations	AFR3	0.815	0.664	0.327	0.814	
	AFR4	0.748	0.560			

FL = Factor Loading, FLS = Factor Loading Squared, AVE= Average Variance Extracted, CR= Composite Reliabilit

Table 4 shows that the loadings of all items range from 0.543 to 0.841. According to Bollen (2014), a recommended factor loading of 0.50 or higher is acceptable, while a loading of 0.70 or higher is considered ideal. Therefore, the results are accepted.

Convergent validity can be evaluated using factor loadings, composite reliability (CR), and average variance extracted (AVE). According to (Hair, Ringle, & Sarstedt, 2011)The results indicate that composite reliability values range from 0.765 to 0.912, exceeding 0.7, which reflects good internal consistency. Additionally, the average variance extracted (AVE) values ranged from 0.522 to 0.668, which are above the cut-off value of 0.50, justifying the use of the construct. Thus, all latent variables met the criteria for establishing convergent validity.

	<u> </u>				
	LI	TI	SY	TQ	AFR
LI					
TI	0.579				
SY	0.3	0.36			
TQ	0.174	0.237	0.799		
AFR	0.575	0.84	0.348	0.126	

Table 5: HTMT Analysis

Table (5) demonstrates that all HTMT values are less than 0.85, which indicates that there are no issues with discriminant validity. According to Henseler, Ringle, and Sarstedt (2015), HTMT values below 0.90 establish discriminant validity between reflective constructs. Consequently, the findings indicate an absence of collinearity problems among the latent constructs (multicollinearity) and reveal no overlapping items in the respondents' perceptions across the evaluated constructs. Page **7**

3.6. Research Hypotheses

Research hypotheses are untested statements that propose a relationship or association between two or more variables. They are formulated based on the study's research question and the existing body of literature. Research hypotheses are crucial for several reasons. Firstly, they provide a clear direction for the research by focusing on particular factors and relationships. Secondly, they allow researchers to evaluate their hypotheses and assess the validity of their proposed explanations, which is crucial for expanding scientific knowledge and highlighting areas that require further investigation. Lastly, hypotheses can help construct a framework for data analysis by guiding researchers on the appropriate statistical tests and techniques relevant to their studies.

Developing a research hypothesis involves formulating a tentative explanation or prediction about the relationship between variables. In this study, a preliminary hypothesis was developed and subsequently refined and modified based on further research and data analysis.

This study proposes nine hypotheses:

Hypothesis 1: The demographic data, including gender, age, education, and English proficiency, have affected the participants' attitudes and future recommendations.

Hypothesis 2: The linguistic issues have affected the attitudes and future recommendations of the participants.

Hypothesis 3: The technical aspects have affected the attitudes and future recommendations of the participants.

Hypothesis 4: Translation quality has affected the attitudes and future recommendations of the participants.

Hypothesis 5: Synchronisation has affected the attitudes and future recommendations of the participants. **Hypothesis 6:** There is a significant difference between Dübverse and ElevenLabs in terms of handling linguistic subtleties and idiomatic expressions.

Hypothesis 7: There is a significant difference between Dübverse and ElevenLabs in technical performance. Hypothesis 8: There is a significant difference in translation quality between Dübverse and ElevenLabs. Hypothesis 9: There is a significant difference between Dübverse and ElevenLabs in terms of synchronisation.

3.7. Study Procedures

The procedures that are followed quantitatively:

- A questionnaire is designed using Google Forms.
- The questionnaire consists of two main sections. The first section includes demographic information related to the participant's gender, age, and English proficiency, which is collected. The second section aims to investigate viewers' reactions to the various genres of Arabic dubbed videos.
- To test the questionnaire's validity, it is examined by three experts in the field for their comments.
- To test the questionnaire's reliability, the responses of 20 participants are analysed statistically using Cronbach's Alpha reliability test.
- The participants' responses will be statistically analysed.

4. Analysis and Findings

This section examined the reactions of participants toward the AI-generated Arabic dubbed videos from ElevenLabs and Dübverse websites. The researchers began with a pilot study involving 25 participants, whose responses were analysed to assess their experience of watching an English video alongside its AI-generated Arabic dubbed versions. The researchers proceeded with data collection, expanding the sample size to 150 participants.

4.1. Sample Characteristics (Demographic Data)

In this section, the researchers analyse the responses of the 150 participants. The analysis first covers the descriptive statistics for the demographic variables, as discussed in Table 6, followed by the responses to the 19 questionnaire items, as presented in Table 7.

Variable	Category	Counts	Per cent
	Males	30	20%
Gender	Females	120	80%
	Total	150	100%
	18-24	65	43.3%
A	25-30	74	49.3%
Age	30+	11	7.3%
	Total	150	100%
	BA/ BSc	85	56.7%
Education	MA or above	65	43.3%
	Total	150	100%
	Intermediate	37	24.7%
English Proficiency	Upper-intermediate	60	40%
	Advanced	53	35.3%
	Total	150	100%

Table 6: Descriptive statistics for the demographic variables

Table 6 indicates that 80% of the participants are female, while 20% are male. It also reveals that nearly half (49.3%) of the participants are between 25 and 30 years old, 43.3% are between 18 and 24 years old, and 7.3% are over 30 years old. Additionally, 56.7% of the participants hold a Bachelor's degree, while 43.3% have earned a Master's degree or higher. Furthermore, the results indicate that 24.7% of the participants have intermediate English proficiency, 40% possess upper-intermediate English proficiency and 35.3% demonstrate advanced proficiency in English.

4.2. Analysis of the Questionnaire's Items

The researchers compared the participants' responses regarding AI-generated Arabic dubbed videos, which were created using AI-driven dubbing websites, namely ElevenLabs and Dübverse, as shown in Table 7. The "% Agree" column indicates the percentage of participants who agreed with the provided statements, while the "% disagree" column represents the percentage of those who disagreed.

Construct	No	Itom	Dubbing	% Dis-	%	%
Construct	110.	Item	EmDubbing% Disagree% Uncern the AI-dubbed vere e for the context andDübverse 58.7 28 e for the context andDübverse 58.7 28 e and grammar of the unded natural.Dübverse 72 16 unded natural.Dübverse 72 16 unded natural.Dübverse 72 16 version.Dübverse 72 16 ubbed videos are clear.Dübverse 46 26 .n or melody) was sat-Dübverse 73.3 10 .ElevenLabs 6 12 11.3 39 .n the dubbed dialogue ressed.Dübverse 76.7 10 .ElevenLabs 14.7 14 14.7 20 .of speech) was accept- pause coincided with an actor to utter them.Dübverse 67.3 17 .ElevenLabs 15.3 37 . 37 . 36 .res was in consonance te translation.Dübverse 60 20 Dübverse 60 20 24 34 .d lip movements were te translation.Dübverse 67.3 12 .original dialogue was in the AI-dubbed ver-Dübverse 67.3 12 .ElevenLabs 67.3 12 . 14	Uncertain	Agree	
		The lexical choices in the AI-dubbed ver-	Dübverse	58.7	28	13.3
	1	sion were appropriate for the context and genre of the video.	ElevenLabs	6	16.7	77.3
Linguistic	2	The sentence structure and grammar of the	Dübverse	72	16	12
IssueS	2	AI-dubbed version sounded natural.	ElevenLabs	13.3	44	42.7
		Different speaking styles and registers (e.g.,	Dübverse	46	26.7	27.3
	3	formal vs. informal speech) were consid- ered in the AI-dubbed version.	ElevenLabs	11.3	39.3	49.3
	4		Dübverse	73.3	10.7	16
	4	The voices in the AI-dubbed videos are clear.	ElevenLabs	6	12	82
	5	The intonation (pitch or melody) was sat-	Dübverse	76.7	10.7	12.7
	3	isfactory.	ElevenLabs	14.7	14	71.3
	6	The accent and tone in the dubbed dialogue	Dübverse	64.7	17.3	18
Technical	6	were adequately addressed.	ElevenLabs	14.7	20.7	64.7
Issues	7	The rhythm (pacing of speech) was accept-	Dübverse	67.3	17.3	15.3
	/	able.	ElevenLabs	12.7	24.7	62.7
	8	Voice selection in terms of dialect, voice	Dübverse	72.7	8	19.3
		cloning, and gender.	ElevenLabs	15.3	37.3	47.3
		Sound effects and background noise were	Dübverse	64	17.3	18.7
	ĺ	taken care of.	ElevenLabs	46.7	19.3	34
	10	The timing of each character's speech, every sentence, phrase, and pause coincided with	Dübverse	74.7	13.3	12
		the timing needed by an actor to utter them.	ElevenLabs	32.7	36.7	30.7
Synchroni-	11	The meaning of gestures was in consonance	Dübverse	60	20	20
sation	11	with the translation.	ElevenLabs	24	34.7	41.3
	12	The speaker's face and lip movements were	Dübverse	82	6.7	11.3
		in consonance with the translation.	ElevenLabs	64.7	16	19.3
	12	The meaning of the original dialogue was	Dübverse	67.3	12.7	20
	15	sion.	ElevenLabs	6.7	11.3	82
Translation	14	The lexical choice of the AI-dubbed version	Dübverse	65.3	16.7	18
Quality	17	was appropriate.	ElevenLabs	6	40.7	53.3
			Dübverse	60	21.3	18.7
	15	Iranslating culture-bound expressions into was appropriate.	Eleven- LabS	9.3	36.7	54

Table 7: The responses of the participants to the questionnaire's item

Attitudes	16	The experience of watching dubbed materials using AI tools is engaging.	54.7	14.7	30.7
	17	AI dubbing is useful for making content accessible to a wider audience.	8	12	80
Recommen-	18	AI- dubbing should consider standard and vernacular accents.	4	4.7	91.3
dations	19	Governments and regulatory bodies should implement guidelines for using AI dubbing, especially ethical con- siderations.	2	4	94

The first three items of the questionnaire focused on gathering participants' feedback on linguistic issues related to AI Arabic dubbed videos generated using the ElevenLabs and Dübverse websites. In the case of Dübverse, 58.7% of the participants disagreed that the lexical choices in the AI-dubbed version were appropriate for the context and genre of the video. Conversely, 77.3% of the participants agreed that the lexical choices in ElevenLabs's AI-dubbed version were suitable for the video's context and genre.

Responses to item 2 addressed whether the sentence structure and grammar of the AI-dubbed version sounded natural. For Dübverse, 72% of the participants disagreed with the statement, indicating that they didn't find the sentence structure and grammar to be natural. In contrast, for ElevenLabs, 42.7% of the participants agreed with the statement, suggesting that they perceived the sentence structure and grammar as more natural.

Responses to item 3 indicated that the AI-generated versions took into account different speaking styles and registers (e.g., formal vs. informal speech) among the participants. Specifically, 46% disagreed with the statement for the case of Dübverse, while 49.3% agreed with it for the case of ElevenLabs.

Items 4 to 9 were designed to gather responses about participants' reactions to the technical issues related to AI Arabic dubbed videos generated using ElevenLabs and Dübverse websites. Specifically, the feedback aimed to evaluate the clarity of the voice in the AI-dubbed videos, the satisfaction of the intonation (pitch or melody), and whether the accent and tone of the dubbed dialogue were adequately addressed. Additionally, it assessed whether the rhythm (i.e., the pacing of speech) was deemed acceptable.

Regarding item 4, which focused on the clarity of the voices in the AI Arabic-dubbed videos, 73.3% of the participants disagreed that the voices in the AI-dubbed version generated by Dübverse were clear. In contrast, 82% of the participants agreed that the voices in the AI-dubbed version generated by ElevenLabs were clear.

For item 5, which evaluated satisfaction with intonation (pitch or melody), 76.7% of the participants disagreed that the intonation was satisfactory for Dübverse, while 71.3% agreed with the statement for ElevenLabs.

In item 6, concerning the adequacy of the accent and tone in the dubbed dialogue, 64.7% of the participants expressed that Dübverse did not adequately address these aspects. In contrast, the same percentage, 64.7%, presented that ElevenLabs addressed them adequately. Yingmin (2023, p. 26) stated that "The purpose of language art is to convey the voice of the heart, i.e., to convey emotion." The participants in the questionnaire were sensitive to this. Many of them felt that this lack of emotional depth affected their overall experience.

In item 7, which assessed whether the rhythm (pacing of speech) was acceptable, 67.3% of the participants disagreed that the rhythm was acceptable for Dübverse, while 62.7% agreed that it was acceptable for ElevenLabs.

Regarding item 8, concerning the voice selection in terms of dialect, voice cloning, and gender, 72.7% of the participants expressed that the voice selection in terms of dialect, voice cloning, and gender was inappropriate for Dübverse. In comparison, 47.3% expressed that it was appropriate for ElevenLabs.

In response to item 9 regarding sound effects and background noise, 64% of the participants disagreed with the statement in the case of Dübverse, while 46.7% disagreed in the case of ElevenLabs.

Items 10 to 12 focused on participants' reactions to the synchronisation of the AI-dubbed videos. For item 10, 74.7% of the participants disagreed that Dübverse ensured that the timing of each character's speech, every sentence, phrase, and pause aligned with the timing required for an actor to utter them. In contrast, only

32.7% of the participants disagreed regarding ElevenLabs.

Regarding item 11, 60% of the participants disagreed that the meaning of gestures was consistent with the translation in the case of Dübverse. In comparison, 41.3% of the participants agreed that the meaning of gestures aligned with the translation for ElevenLabs.

Item 12 of the questionnaire was intended to gather responses regarding whether the speaker's face and lip movements aligned with the translation. 82% of the participants disagreed with the statement for Dübverse, whereas 64.7% of the participants disagreed with the statement for ElevenLabs. This is inconsistent with some scholars (see Chaume, 2020; Yuan & Jin, 2023), highlighting the importance of synchronisation, that the utterances should be aligned with the lip movements to create more natural dubbing.

Items 13 to 15 of the questionnaire focused on gathering responses about participants' reactions to the technical quality of the AI-driven dubbing websites. The feedback aimed to determine whether AI-dubbed versions accurately conveyed the meaning of the original dialogue, made appropriate lexical choices of the AI-dubbed versions, or effectively translated culture-bound expressions. In response to item 13, which asked whether the original dialogue's meaning was accurately conveyed in the AI-dubbed versions, 67.3% of the participants disagreed regarding Dübverse, while 82% agreed that ElevenLabs accurately conveyed the original dialogue's meaning.

In responses to item 14 regarding whether the lexical choice of the AI-dubbed versions was appropriate, 65.3% of the participants expressed that the linguistic choice of the AI-dubbed version was inappropriate for Dübverse. In comparison, 53.3% expressed that the lexical choice of the AI-dubbed version was appropriate for ElevenLabs.

In response to item 15, which assessed the appropriateness of translating culture-bound expressions in the AI-dubbed versions, 60% of the participants disagreed with the statement for Dübverse. In comparison, 54% of the participants agreed that ElevenLabs handled these expressions appropriately. This lines with the findings of Brannon et al. (2023) study, stating that dubbing faces sociocultural constraints, particularly in translating culture-bound expressions, which must be adapted to align with the linguistic and cultural context.

Items 16 to 19 were designed to gather insights regarding participants' Attitudes and Future Recommendations concerning AI dubbing. The feedback aims to determine whether participants find the experience of watching dubbed materials using AI tools engaging, whether AI-dubbing is effectively enhancing the content accessibility to a broader audience, also whether the implementation of AI-dubbing should consider both the standard and vernacular accents and if governments and regulatory bodies should establish guidelines for the ethical considerations of using AI-dubbing.

The responses revealed that 54.7% of the participants disagreed that watching dubbed materials using AI tools is an engaging experience. Conversely, 80% of the respondents acknowledged that AI-dubbing is beneficial for making content more accessible to a broader audience. Additionally, 91.3% of the participants asserted the importance of considering both standard and vernacular accents in AI dubbing, and 94% agreed that governments and regulatory bodies should establish guidelines regarding ethical considerations of using AI in dubbing.

4.3. Testing the Hypotheses

This section tests the nine hypotheses outlined in the methodology section.

Hypothesis 1

Alternative Hypothesis (H1): The demographic data, including gender, age, education, and English proficiency, have affected the participants' attitudes and future recommendations.

Null Hypothesis (H01): The demographic data, including gender, age, education, and English proficiency, have not affected the participants' attitudes and future recommendations.

1			0 0	0				
			Estimate	SE.	CR.	Р	Effect	R2
Gender	\rightarrow	AFR	-0.069	0.079	-0.869	0.385	-0.071	
Age	\rightarrow	AFR	0.036	0.052	0.699	0.485	0.057	0.012
Education	\rightarrow	AFR	-0.04	0.064	-0.63	0.528	-0.051	0.012
English Proficiency	\rightarrow	AFR	-0.015	0.041	-0.363	0.716	-0.03	

Table 8: Structural Equation Modelling Regression weights

S.E. = Standard errors of the regression weights, C.R. = Critical Ratio, P = p-value (*<0.05, **<0.01, ***<0.001)

Table (8) shows no statistically significant effects for the demographic data, including (gender, age, education, and English proficiency) on the participants' attitudes and future recommendations regarding AI Arabic dubbed videos generated using ElevenLabs and Dübverse websites. As indicated by the regression weights, the p-values (0.385, 0.485, 0.528, and 0.716)) respectively, are more significant than 0.05, and the critical ratio values are less than 2 (Byrne, 2013). Consequently, the first H0 was not rejected, indicating that there is no significant evidence that demographic data affects audience attitudes.

In addition, the demographic data account for just 1.2% of the variation in attitudes and future recommendations regarding AI-generated videos, with an R-squared value of 0.012.

Hypothesis 2

H2: The linguistic issues have affected the attitudes and future recommendations of the participants.

H02: The linguistic issues have not affected the attitudes and future recommendations of the participants.

			Estimate	SE.	CR.	Р	Effect	R2
Linguistic Issues	\rightarrow	AFR	0.255	0.070	3.632	* * *	0.285	0.081

Table 9: Structural Equation Modelling Regression weights

S.E. = Standard errors of the regression weights, C.R. = Critical Ratio, P = p-value (*<0.05, **<0.01, ***<0.001)

Table 9 shows that linguistic issues statistically affect the participants' attitudes and future recommendations regarding AI-generated Arabic dubbed videos using ElevenLabs and Dübverse websites. The regression weights indicate that the p-value (***) is less than 0.0.01, and the critical ratio value is greater than 2 (Byrne, 2013). Accordingly, H2 is accepted, indicating that the linguistic issues have affected the attitudes and future recommendations of the participants. This is consistent with Yingmin (2023), who addresses the linguistic issues, particularly in AI voiceovers, focusing on how AI struggles to replicate the natural flow of speech and the linguistic tensions found in human speech, especially in the Mandarin language.

In addition, linguistic issues account for 8.1% of the variation in attitudes and future recommendations regarding AI-dubbed videos, with an R-squared equaling 0.081. The effect size of linguistic issues on AFR equals 0.285.

Hypothesis 3

H3: The technical aspects have affected the attitudes and future recommendations of the participants.

H03: The technical aspects have not affected the attitudes and future recommendations of the participants.

Table 10: Structural Equation Modelling Regression weights

			Estimate	SE.	CR.	Р	Effect	R2
Technical Issues	\rightarrow	AFR	0.353	0.068	5.199	* * *	0.392	0.154

S.E. = Standard errors of the regression weights, C.R. = Critical Ratio, P = p-value (*<0.05, **<0.01, ***<0.001)

Table (10) shows a statistically significant effect of technical issues on the participants' attitudes and future recommendations regarding AI Arabic dubbed videos generated using ElevenLabs and Dübverse websites. The regression weights indicate that the p-value (***) is less than 0.0.01, and the critical ratio value is greater than 2 (Byrne, 2013). Accordingly, H3 is accepted, indicating that the technical issues have affected the attitudes and future recommendations of the participants.

In addition, technical issues account for 15.4% of the variation in attitudes and future recommendations regarding AI-dubbed videos, with an R-squared value of 0.154. The effect size of (Technical Issues) on (AFR) equals 0.392. This is in line with the findings of Abulawi et al. (2022), which showed how technical issues have affected participants' attitudes and future recommendations. Moreover, the study of Baños (2023)aligns with this research, as it highlighted that the AppTek dubbing system acknowledges a limitation in the synthesised voice's flat tone, which fails to capture the rhythm and intonation of the original video.

Hypothesis 4

H4: Translation quality has affected the attitudes and future recommendations of the participants.

H04: Translation quality has affected the attitudes and future recommendations of the participants.

Table 11: Structural Equation Modelling Regression weights

			Estimate	SE.	CR.	Р	Effect	R2	
Translation quality	\rightarrow	AFR	0.219	0.066	3.328	* * *	0.263	0.069	
S.E. = Standard errors of the r	egress	ion weig	hts, $C.R. = C$	ritical Rati	o, P = p-va	lue (*<0	.05, **<(0.01, ***<().001)

Table 11 shows that there is a statistically significant effect of translation quality on the participants' attitudes and future recommendations regarding AI-generated Arabic dubbed videos using ElevenLabs and Dübverse websites. The regression weights indicate that the p-value (***) is less than 0.0.01, and the critical ratio value is greater than 2 (Byrne, 2013). Accordingly, the fourth alternative hypothesis is accepted, indicating that the

In addition, the quality of translation accounts for 6.9% of the variation in attitudes and future recommendations regarding AI-dubbed videos, with an R-squared of 0.069. The effect size of translation quality on AFR equals 0.263.

Hypothesis 5

H5: Synchronisation has affected the attitudes and future recommendations of the participants.

translation quality has an impact on the attitudes and future recommendations of the participants.

H05: Synchronisation has not affected the attitudes and future recommendations of the participants.

EstimateSE.CR.PEffectR2Synchronisation \rightarrow AFR0.2040.0563.619***0.2840.081

 Table 12:
 Structural Equation Modelling Regression weights

S.E. = Standard errors of the regression weights, C.R. = Critical Ratio, P = p-value (*<0.05, **<0.01, ***<0.001)

Table (12) demonstrates a statistically significant effect of synchronisation on the participants' attitudes and future recommendations regarding AI Arabic dubbed videos generated using ElevenLabs and Dübverse websites. As shown by the regression weights, the p-value (***) is less than 0.0.01, and the critical ratio value is greater than 2 (Byrne, 2013). Accordingly, the fifth alternative hypothesis is accepted, indicating that synchronisation has affected the attitudes and future recommendations of the participants.

Moreover, synchronisation accounts for 8.1% of the variation in attitudes and future recommendations concerning AI-dubbed videos, with an R-squared value of 0.081. The effect size of (Synchronisation) on (AFR) equals 0.284.

Hypothesis 6

H6: There is a significant difference between Dübverse and ElevenLabs in terms of handling linguistic subtleties and idiomatic expressions.

H06: There is not a significant difference between Dübverse and ElevenLabs in terms of handling linguistic subtleties and idiomatic expressions.

Table 13: Independent Samples Test results

	Category	Mean (SE.)	T-value	Sig. t
Handling linguistic subtleties and idiomatic expressions	Dübverse	2.57 (0.05)	13.807	0.000
	ElevenLabs	3.46 (0.04)		

S.E. (Standard error of MEAN, ***p < 0.001, **p < 0.01, *p < 0.05

Table (13) shows a statistically significant difference between Dübverse and ElevenLabs in terms of handling linguistic subtleties and idiomatic expressions. The significant level (sig. t) value was less than (0.001), and the Mean values show that ElevenLabs has better handling of linguistic subtleties and idiomatic expressions than Dübverse.

Hypothesis 7

H7: There is a significant difference between Dübverse and ElevenLabs in technical performance.

H07: There is no significant difference between Dübverse and ElevenLabs in technical performance.

 Table 14: Independent Samples Test results

	Category	Mean (SE.)	T-value	Sig. t	
Technical Performance	Dübverse	2.47 (0.05)	14.937	0.000	
	ElevenLabs	3.42 (0.04)			
$S = \sqrt{S_{\text{transformation}}} = 0.001 \times 10^{-1} = 0.001$					

S.E. (Standard error of MEAN, ***p < 0.001, **p < 0.01, *p < 0.05

Table (14) shows a statistically significant difference between Dübverse and ElevenLabs in technical performance. The significant level (sig. t) value was less than (0.001). The Mean values show that ElevenLabs has better technical performance than Dübverse. So, we reject the null hypothesis in favour of supporting H7.

Hypothesis 8

H8: There is a significant difference in translation quality between Dübverse and ElevenLabs.

H08: There is not a significant difference in translation quality between Dübverse and ElevenLabs.

 Table 15: Independent Samples Test results

	Category	Mean (SE.)	T-value	Sig. t
Translation Quality	Dübverse	2.55 (0.06)	14.002	0.000
	ElevenLabs	3.56 (0.04)	14.092	

S.E. (Standard error of MEAN, ***p < 0.001, **p < 0.01, *p < 0.05)

Table (15) shows a statistically significant difference between Dübverse and ElevenLabs in translation quality. The significant level (sig. t) value was less than (0.001). The Mean values show that ElevenLabs has better translation quality than Dübverse. This means that translation quality indeed varies between the two websites.

Hypothesis 9

H9: There is a significant difference between Dübverse and ElevenLabs in terms of synchronisation.

H09: There is not a significant difference between Dübverse and ElevenLabs in terms of synchronisation.

	Category	Mean (SE.)	T-value	Sig. t
Synchronisation -	Dübverse	2.42 (0.06)	((01	0.000
	ElevenLabs	2.90 (0.06)	6.681	
S.E. (Standard error of MEAN, ***p < 0.001, **p < 0.01, *p < 0.05				

Table 16: Independent Samples Test results

Table (16) shows a statistically significant difference between Dübverse and ElevenLabs in terms of synchronisation. The significance level (sig. t) value was less than 0.001. The Mean values show that ElevenLabs significantly exceeds Dübverse in terms of synchronisation.

5. Conclusion

Based on the analysis, testing the research hypotheses has shown that all four variables comprising the first section of the questionnaire, precisely the demographic data, including gender, age, education, and English proficiency, were found to show statistically insignificant differences between the participants' attitudes and future recommendations regarding AI Arabic dubbed videos using AI-driven dubbing websites.

The linguistic issues have significantly impacted the participants' attitudes and future recommendations regarding AI-generated Arabic dubbed videos from ElevenLabs and Dübverse websites. The researchers computed the subscale Means for each construct and conducted a Regression test. The p-value was equal to 0.000 (less than 0.001), indicating a statistically significant effect of the independent variable "Linguistic Issues" on the participants' attitudes and future recommendations regarding AI Arabic dubbed videos generated using ElevenLabs and Dübverse websites. The Beta values showed that Linguistic Issues account for a 28.5% effect size on the participants' attitudes toward the AI-dubbed videos.

The technical issues have significantly impacted the participants' attitudes and future recommendations regarding AI-generated Arabic dubbed videos from ElevenLabs and Dübverse websites. The researchers calculated the subscale Means for each construct and conducted a Regression test. The p-value was 0.000 (less than 0.001), indicating a statistically significant effect of the independent variable "Technical Issues" on the participants' attitudes and future recommendations regarding AI-dubbed videos. The Beta values further demonstrated that the effect size of "Technical Issues" accounts for 39.2% of the variation in the participants' attitudes toward the AI-dubbed videos.

The translation quality significantly affected the participants' attitudes and future recommendations regarding AI-generated Arabic dubbed videos from Elevenlabs and Dübverse websites. The researchers computed the subscale Means for each construct and conducted a Regression test. The p-value was equal to 0.000 (less than 0.001), indicating a statistically significant effect of the independent variable "Translation Quality" on the participants' attitudes and future recommendations regarding AI-dubbed videos. The Beta values revealed that the effect size of "Translation Quality" accounts for 26.3% of the participants' attitudes toward AI-dubbed videos.

Synchronisation significantly affected the participants' attitudes and recommendations regarding AI Arabic dubbed videos generated using Elevenlabs and Dübverse websites. The researchers calculated the subscale Means for each construct and conducted a Regression test. The p-value was equal to 0.000 (less than 0.001), indicating a statistically significant effect of the independent variable "Synchronisation" on the participants' attitudes and future recommendations regarding AI-dubbed videos. The Beta values revealed that the effect size of "Synchronisation" accounts for 28.4% of the participants' attitudes toward AI-dubbed videos.

A statistically significant difference is identified between Dübverse and ElevenLabs in terms of handling linguistic subtleties and idiomatic expressions. The researchers computed the subscale Means for each construct and conducted an Independent Sample T-Test. ElevenLabs has better handling of linguistic subtleties and Page **16**

idiomatic expressions than Dübverse.

A statistically significant difference in technical performance is identified between Dübverse and ElevenLabs. The researchers computed the subscale Means for each construct and conducted an Independent Sample T-Test. ElevenLabs has better technical performance than Dübverse.

A statistically significant difference is identified between Dübverse and ElevenLabs in translation quality. The researchers computed the subscale Means for each construct and conducted an Independent Sample T-Test. ElevenLabs has better translation quality than Dübverse.

A statistically significant difference is identified between Dübverse and ElevenLabs in terms of synchronisation. The researchers computed the subscale Means for each construct and conducted an Independent Sample T-Test. ElevenLabs is better in terms of synchronisation than Dübvers.

Funding: This research received grant no. (320/2024) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

Acknowledgments: The authors would like to thank the reviewers for their comments.

Conflicts of Interest: The authors declare no conflict of interest.

Biographies

Maisaa Ahmad is a Master's holder in Audio-visual and Mass Media Translation at the Applied Science Private University. Her main areas of interest include discourse analysis and translation studies.

Ahmad S Haider (Corresponding) is an associate professor in the Department of English Language and Translation at the Applied Science Private University, Amman, Jordan. He received his Ph.D. in Linguistics from the University of Canterbury/ New Zealand. His current research focuses on how political events are socially, discursively, and linguistically represented in media combining Corpus Linguistics and (Critical) Discourse Analysis. His main areas of interest include corpus linguistics, discourse analysis, pragmatics, and translation studies. Dr. Haider has built different large Arabic and English corpora. He professionally masters different Corpus Linguistic software packages. http://orcid.org/0000-0002-7763-201X

Hadeel Saed is a Professor at the Department of English Language and Translation, Applied Science Private University (ASU). Dr. Saed received her PhD in Curriculum and Methods of Teaching English from the university of Jordan 2012. Her main areas of interest include second language learning, first language learning, translation studies, socio-linguistics, syntax, and phonetics. Dr. Saed has published many research articles in prestigious journals. https://orcid.org/0000-0001-6833-2165

Authorship and Level of Contribution

Conceptualization, M.A., A.S.H., and H.S.; methodology, A.S.H.; validation, M.A., A.S.H., and H.S.; formal analysis, A.S.H.; investigation, M.A. and H.S.; resources, M.A. and H.S.; writing—original draft preparation, M.A. and H.S.; writing—review and editing, A.S.H.; visualization, H.S.; supervision, A.S.H.; project administration, A.S.H.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

References

- Abulawi, F., Al-Salman, S., & Haider, A. (2022). Modern Standard Arabic vs. Egyptian Vernacular in Dubbing Animated Movies: A Case Study of A Bug's Life. *International Journal of Communication and Linguistic Studies*, 21, 125-141. doi:10.18848/2327-7882/CGP66
- Alrousan, F., & Haider, A. S. (2022). Dubbing television advertisements across cultures and languages: A case study of English and Arabic. *Language Value*, 15(2), 54-80. doi:https://doi.org/10.6035/ languagev.6922
- Baños, R. (2023). Key challenges in using automatic dubbing to translate educational YouTube videos. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, 22, 61–79.
- Bigioi, D., & Corcoran, P. (2023). Multilingual video dubbing—a technology review and current challenges. *Front. Signal Process*, 3, 1-14. doi:10.3389/frsip.2023.1230755
- Bollen, K. A. (2014). Structural equations with latent variables. New Jersey: John Wiley & Sons.
- Brannon, W., Virkar, Y., & Thompson, B. (2023). Dubbing in practice: A large scale study of human localization with insights for automatic dubbing. *Transactions of the Association for Computational Linguistics*, 11, 419-435.
- Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* London Routledge.
- Chaume, F. (2004). Synchronization in dubbing: A translational approach. In P. Orero (Ed.), *Topics in audiovisual translation* (pp. 35-52). Amsterdam, The Netherlands: John Benjamins.
- Chaume, F. (2020). Audiovisual translation: dubbing. London: Routledge.
- Chronopoulou, A., Thompson, B., Mathur, P., Virkar, Y., Lakew, S. M., & Federico, M. (2023). Jointly Optimizing Translations and Speech Timing to Improve Isochrony in Automatic Dubbing. *arXiv* preprint arXiv:.12979.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), 297-334. doi:https://doi.org/10.1007/bf02310555
- Federico, M., Enyedi, R., Barra-Chicote, R., Giri, R., Isik, U., Krishnaswamy, A., & Sawaf, H. (2020). From speech-to-speech translation to automatic dubbing. *arXiv preprint arXiv:*.06785.
- Federico, M., Virkar, Y., Enyedi, R., & Barra-Chicote, R. (2020). *Evaluating and optimizing prosodic alignment for automatic dubbing*. Paper presented at the INTERSPEECH, Shanghai, China.
- George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update. . Boston: Allyn & Bacon.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory* and Practice, 19(2), 139-152.
- Hayes, L. (2021). Netflix Disrupting Dubbing: English Dubs and British Accents. *Journal of Audiovisual Translation*, 4(1), 1-26. doi:https://doi.org/10.47476/jat.v4i1.2021.148
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, 43, 115-135.
- Iitola, M.-M. (2017). "Let It Go". Subtitling and Dubbing Song Lyrics into Finnish in the Animation Film Frozen. (Master's Degree Programme in Language expertise in specialized society). University of Vaasa, Finland Retrieved from http://osuva.uwasa.fi/handle/10024/880
- Lakew, S. M., Federico, M., Wang, Y., Hoang, C., Virkar, Y., Barra-Chicote, R., & Enyedi, R. (2021). Machine translation verbosity control for automatic dubbing. Paper presented at the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Canada.
- Lee, A., Chen, P.-J., Wang, C., Gu, J., Popuri, S., Ma, X., . . . Tang, Y. (2021). Direct speech-to-speech translation with discrete units. *arXiv preprint*.
- Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.
- Öktem, A., Farrús, M., & Bonafonte, A. (2019). Prosodic phrase alignment for machine dubbing. arXiv preprint.
- Trivedi, A., Pant, N., Shah, P., Sonik, S., & Agrawal, S. (2018). Speech to text and text to speech recognition systems-Areview. *IOSR J. Comput. Eng*, 20(2), 36-43.
- Wahlster, W. (2013). Verbmobil: foundations of speech-to-speech translation. Berlin: Springer Science &

Page 19

Business Media.

- Yahiaoui, R. (2016). Ideological constraints in dubbing The Simpsons into Arabic. *Altre Modernità: Rivista di studi letterari e culturali*(1), 182-200.
- Yingmin, S. (2023). Realistic problems and innovative path of AI dubbing in the era of intelligent media. *Int. J. Front. Socio*, *5*, 23-27.
- Yuan, Z., & Jin, H. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film Mulan. *Linguistica Antverpiensia*, New Series–Themes in Translation Studies, 22, 80–94.