ARTICLE

Captioning for Clarity: Mastering Stress with Word-Based vs. Phonetic Video Transcriptions

Shaimaa Mohamed Mahmoud Helal ⓘ *, Hassan Saleh Mahdi* * ⓘ *, Mustafa Abdulmuhsen Thabet* ⓘ

Faculty of Language Studies, Arab Open University, Riaydh, Saudi Arabia

ABSTRACT

This study explores an understudied aspect of video captioning in language learning: the use of phonetic transcription. While extensive research has examined the effects of traditional word-based captioning, the potential benefits of phonetic transcription captions remain largely unexplored. To address this knowledge gap, a quasi-experimental study was conducted comparing two captioning conditions: traditional word-based captions and phonetic transcription captions. Two groups of English as a Foreign Language (EFL) learners were purposively selected based on their course enrollment, with participants volunteering for the study. The control group viewed videos with standard word captions, while the experimental group watched videos captioned with phonetic transcriptions. Pre- and post-tests were administered to assess improvements in word stress pronunciation. The results revealed no statistically significant difference between the post-test scores of the control and experimental groups. This suggests that both captioning methods - words and phonetic transcription - positively impact stress pronunciation learning. The findings indicate that enhanced stress pronunciation may contribute to overall language intelligibility and fluency, potentially leading to improved language learning outcomes. This study opens avenues for future research on the application of phonetic transcription in other areas of pronunciation, such as intonation and connected speech.

Keywords: EFL learners; Pronunciation; Stress; Video captioning

*CORRESPONDING AUTHOR:

Hassan Saleh Mahdi,Faculty of Language Studies, Arab Open University, Riaydh, Saudi Arabia;Email: h.mahdi@arabou.edu.sa

1. Introduction

Video captioning serves as an effective tool that holds significant potential for enhancing language acquisition (Hsieh, 2020; Teng, 2023). Its mechanism involves the simultaneous display of textual content alongside video material, equipping learners with a written aid corresponding to the spoken discourse. This feature proves particularly advantageous for learners who encounter unfamiliar lexical items in a new language. Video captioning is defined as "the translation of acoustic sounds into text, and that is how and why it could foster listening comprehension" (Hsieh, 2020, p. 1). Poviding learners with a visual representation of spoken content, especially when the material surpasses their proficiency level, facilitates the identification and comprehension of meaningful linguistic segments during video consumption (Avello & Muñoz, 2023).

While the effectiveness of video captioning in language learning has been well-documented, a significant gap exists in our understanding of how different types of captioning might impact specific aspects of language acquisition, such as pronunciation and stress patterns. Video captions can be presented in two main ways: spelling (using the standard written form of the language) or phonetic transcription (employing symbols to represent the actual sounds of spoken words). Previous research (e.g., Perez, et al., 2014; Winke, et al., 2010) has primarily focused on the effectiveness of video captions with spelled words. However, the potential of phonetic transcription in video captioning remains largely unexplored, particularly in its application to learning English stress patterns.

This study aims to bridge this gap by examining how video captions with phonetic transcription can specifically impact learners' ability to master English stress patterns. We assume that phonetic transcription, by providing a more accurate representation of pronunciation, may offer unique benefits in this area. To prove this assumption, the study was designed to compare the effectiveness of traditional word-based captions versus phonetic transcription-based captions on EFL learners' pronunciation of English stress. The study employed a sample of forty EFL undergraduates at Arab Open University, Saudi Arabia. The significance of this research lies in its potential to inform pedagogical practices in language teaching, particularly in the field of pronunciation instruction. If phonetic transcription proves to be more effective in teaching stress patterns, it could lead to the development of new educational materials and technologies that incorporate this method. Furthermore, this study contributes to the broader field of computer-assisted language learning (CALL) by exploring innovative ways to leverage technology for specific language learning goals.

The implications of this research extend beyond the classroom. Improved stress pronunciation can enhance overall intelligibility and fluency, potentially leading to better outcomes in learning a language. Moreover, this study may pave the way for future studies exploring the use of phonetic transcription in other areas of language learning, such as intonation or connected speech phenomena.

2. Literature review

2.1 Teaching pronunciation

Effective pronunciation instruction plays a pivotal role in the journey of foreign language learners, serving as a cornerstone that not only facilitates comprehension but also ensures clear and effective communication, as highlighted by Dao et al. (2021). It serves as a bridge between EFL learners and native speakers, enabling them to navigate the intricacies of pronunciation with confidence and accuracy. Recognizing the significance of this, teachers are increasingly adopting a multi-faceted approach, drawing upon insights from research and pedagogy to design comprehensive instructional strategies. This approach, as advocated by Darcy and Rocca (2022), encompasses a spectrum of techniques aimed at addressing various aspects of pronunciation. Explicit instruction on individual sounds and suprasegmentals, such as stress and intonation, forms the bedrock of this approach, providing learners with a solid foundation upon which to build their pronunciation skills. Furthermore, it incorporates many opportunities for immersive listening and speaking practice, allowing learners to internalize and apply what they have learned in authentic communicative contexts. Through this comprehensive approach, teachers can empower learners to not only produce sounds accurately but also to grasp the differences in rhythm, melody, and stress patterns inherent in the target language, thereby enhancing their overall communicative competence. By fostering a supportive learning environment and offering targeted feedback, educators can guide learners on their pronunciation journey and equip

them with the necessary skills to engage effectively in oral communication and navigate the complexities of real-world interactions.

2.2 Technology and pronunciation

The integration of modern technological platforms such as YouTube into educational settings has emerged as a powerful tool to invigorate teaching and learning processes, eliciting heightened student engagement and motivation. The integration of YouTube into educational settings has shown promising results in enhancing teaching and learning processes. Studies have demonstrated that incorporating YouTube videos in classrooms can lead to increased student motivation, improved content learning, and more successful learning outcomes (Alwehaibi, 2015; Tohe, 2021). YouTube's effectiveness as an educational tool is attributed to its ability to provide rich visual resources, engage the Internet generation of students, and create collaborative learning spaces (Duffy & Duffy, 2008). Research has shown that YouTube can be particularly beneficial in language learning, helping to develop students' skills and creativity while making the educational environment more exciting and encouraging (Abbas & Qassim, 2020). However, it is crucial to note that for YouTube's integration to be truly transformative, it must be based on clear philosophical and pedagogical foundations, rather than simply serving as a tool for capitalist consumerism (Tohe, 2021).

2.3 Video captioning and teaching pronunciation

Several studies (e.g., Mahdi, 2017; Mohsen & Mahdi, 2021; Yeldham, 2018) have explored the use of video captioning and computer-assisted pronunciation teaching in language learning. Video captioning involves describing videos with language, benefiting various applications like accessibility and content generation (Cengiz, 2023; Vázquez, 2022). In addition, research on computer-assisted pronunciation teaching has shown positive outcomes, especially with the use of technologies like Automatic Speech Recognition (ASR) programs and YouTube videos (Abdar, et al., 2023; Maspufah, 2022). These studies highlight the importance of technology in enhancing language learning experiences, with a focus on improving pronunciation skills through innovative methods. The findings emphasize the role of teachers in guiding

learners effectively in utilizing these tools for better learning outcomes.

2.4 Video captioning and teaching pronunciation

Video captioning, or "same-language subtitles" (Gernsbacher, 2015, p. 195), is a form of input modality. It entails showing written material on a video screen that corresponds to spoken content in the same language. Therefore, it is defined as "on-screen text in a given language combined with a soundtrack in the same language" (Qiao & Yijun, 2023, p. 14785). Displaying captions on videos has been shown to benefit various groups, among them adolescents, college students, and adults, by improving attention, comprehension, and memory related to the video content (Gernsbacher, 2015). It can increase college students' in-class attention (Steinfeld, 1998), boost adolescents' spoken and written vocabulary (Davey& Parkhill; 2012), and enhance second language learners' pronunciation (Mitterer& McQueen, 2009; Wisniewska & Mora, 2020).

The effects of video captioning have been extensively studied and documented in various research papers. More than 100 empirical studies document that captioning a video improves learners' attention to, comprehension of, and memory of the video (Gernsbacher, 2015). It has been found to be an effective instructional method to enhance both the listening and reading comprehension of a second language (Borras & Lafayette, 1994; Danan, 2004; Garza, 1991; Markham & Peter, 2003). Furthermore, it promotes language comprehension (Bird & Williams, 2002). It also facilitates vocabulary acquisition (Plass et al., 1998), reading (Chun & Plass, 1997), and listening comprehension (Chung, 1999; Danan, 1992; Hwang, 2003; Markham, 2001; Markham & McCarthy,2001). It also aids the learning of word meaning (Hsu et al., 2013), the visualization of aural cues, memory trace of words, and the identification of sounds without textual support (Danan, 2004).

Despite the beneficial aspects described above, captioning has limitations. It may not be appropriate for all materials or viewers with advanced language skills (Danan, 2004, p.71). To be more specific, it may be beneficial to beginners if the contents are carefully tailored to their level and include many familiar phrases that can be triggered and reinforced through video captioning (Danan, 2004, p. 71). Thus,

two factors are working together and affecting captioning beneficial effects. These are language proficiency and material adaptation. Furthermore, captions may hinder learners' ability to synchronize reading with online auditory speech processing, particularly at lower levels of competency (Wisniewska & Mora, 2020). Moreover, captions cannot properly compensate for the quick rate of speech and the difficult vocabulary level if the material is too advanced, resulting in poor language gains (Wisniewska & Mora, 2020). Hence, two additional factors are affecting captioning. These are cognitive overload and comprehensible input.

Several studies have extensively investigated video captioning and reached the consensus that it has a potential effect in aiding lexical items acquisition on the part of EFL learners (Chai & Erlam, 2008; Lo, 2024; Markham, 1999; Perez, 2022; Sydorenko, 2010; Winke et al., 2010). Despite this extensive investigation of video captioning beneficial effects on vocabulary acquisition, little is known about the way it could promote learners' pronunciations. Pronunciation is considered as a component of second language vocabulary acquisition (Nation, 2001). Very few studies have examined the impact of different types of video captioning on learners' pronunciation skills (e.g., Mahdi, 2017; Mohsen & Mahdi, 2021; Wisniewska & Mora, 2017; 2020). With the aim of exploring the role of exposure to L2 captioned videos in L2 pronunciation development, Wisniewska and Mora (2017) investigated the integration of auditory and orthographic input in L2 learning during reading dynamic texts in L2-captioned videos. Thirty-eight L1-Catalan/Spanish learners of L2-English participated in the study. The learners' eye movements are recorded during watching short L2-captioned video clips. In addition to that, the learners' L2 segmentation skills and L2 proficiency are assessed. The results of the study showed that exposure to L2-captioned videos has positive implications for L1-Catalan/Spanish learners of L2-English. This was mainly because the learners showed improved skills in integrating auditory and textual input while reading dynamic texts. Furthermore, the study refers to a key factor in achieving modality integration, which is efficient reading. In the study, successful text-sound integration was indicated by the synchronized fixation of selected lexis with their auditory onsets. The value of bimodal input, as subtitles aid in decoding and segmenting speech, was emphasized in the study. This is due to its (bimodal input) role in improving

word identification and lexical access. The study concluded that what might lead to enhancing L2 pronunciation skills and developing accurate L2 phonological representations was extensive exposure to L2-captioned videos.

Several studies examined the effects of video captioning on second language (L2) pronunciation and vocabulary acquisition. For example, Mohsen and Mahdi (2021) found that both full and partial captioning improved word pronunciation compared to no captioning, with partial captioning resulting in lower cognitive load. Wisniewska and Mora (2020) observed benefits in speech segmentation and processing skills from extended exposure to captioned videos, with pronunciation gains occurring when captions were present without explicit focus on phonetic form or when attention was directed to pronunciation without captions. Mahdi (2017) demonstrated that keyword captioning significantly improved L2 pronunciation compared to full captioning, while both captioning modes were equally effective for vocabulary learning. These studies collectively suggest that various forms of video captioning can enhance L2 pronunciation and vocabulary acquisition, with the effectiveness depending on factors such as captioning type, focus of attention, and duration of exposure.

The impact of video captioning on pronunciation learning finds theoretical support in the Dual Coding Theory, a cognitive framework proposed by Paivio (2014). This theory posits that humans process and retain information more effectively when it is presented through multiple sensory channels, particularly visual and auditory. By applying this principle, video captions serve as a powerful tool for language learners, providing them with simultaneous visual reinforcement alongside auditory input. Through the integration of text with spoken language, learners engage in a dual coding process wherein both visual and auditory stimuli contribute to comprehension and memory formation. This dual coding mechanism facilitates deeper encoding of linguistic information, enhancing learners'' ability to recall and reproduce pronunciation accurately. Mayer (2003) further supports this notion by highlighting the role of visual reinforcement in improving learning outcomes. Thus, video captioning not only aids in comprehension but also strengthens memory retention, ultimately contributing to more effective pronunciation learning experiences for language learners.

While both groups improved, it is important to note that

the learners who were exposed to video captioned in spelled words scored slightly higher in the post-test compared to the experimental group in which learners were exposed to video captioned in phonetic transcription. This can be attributed to the notion that learners might be more familiar with traditional spelling compared to phonetic symbols. This familiarity with spelled words could potentially reduce cognitive load, allowing them to focus more on the pronunciation itself (Nation, 2003). Phonetic transcription, while accurate, might require additional mental processing to decode the unfamiliar symbols, thus potentially hindering some learners initially. Moreover, exposure to spelled words in captions could lead to a transfer of existing spelling knowledge to pronunciation. Learners might subconsciously connect the way a word is spelled to its sound, even if the spelling-to-sound relationship is not always consistent in English (Treiman & Kessler, 2023). It is important to acknowledge that the difference in scores observed in this study was relatively small and might not be statistically significant. Further research with a larger sample size and more robust statistical analysis is needed to confirm this trend definitively.

2.5 Stress

Learning pronunciation is not only restricted to the study of the sounds in a language. It also includes the study of the placement of stress and other suprasegmental features (Huwar & Mehawesh, 2015). In addition to that, the meaning in communication relies on both segmental and suprasegmental features (Sharma, 2023). Segmental features include consonants and vowel phonemes, whereas suprasegmentals comprise several types among them stress and intonation. Suprasegmental features are phonetic features which are not the properties of single consonants or vowels (Ladefoged, 2011). Stress is one of the suprasegmental features of expressions (Ladefoged, 2011). It is an extra effort given to a particular syllable or word (Ladefoged & Johnson, 2014). It can be described as a property of a lexeme (Kreidle, 2008). Furthermore, it is an isolable phonological phenomenon, a suprasegmental phenomenon, a property of a string of segments not of an individual segment (Hogg & McCully, 1987). Roach (2009) refers to four factors which help in identification of the stressed syllable: loudness, length, pitch and quality. Thus, the stressed syllable, according to Roach, should be louder than other syllables in disyllabic or poly-

syllabic words, longer with greater pitch, and should include a vowel which is different in quality with other syllables. For Ladefoged and Johnson (2014), a stressed syllable "is usually produced by pushing more air out of the lungs in one syllable relative to others… [therefore, it] has greater respiratory energy than neighbouring unstressed syllables" (p. 119).

Stress plays a vital role in conveying meaning, mood, and emotion in speech. Pronunciation experts have focused on the suprasegmental feature to enhance oral communication for the past three decades (Morley, 1991). Speakers use stress to emphasize important information, and if they fail to do so effectively, misunderstandings can occur (Alkhuli, 2002). Therefore, incorrect stress placement sometimes hinders intelligibility as well as comprehensibility at word and sentence levels (Ghosh & Levis, 2021; Jenkins, 2002). Additionally, incorrect stress assignment, as a result of mispronunciation in which the vowel quality is changed, for Cutler (2015), "is highly likely to impede successful recognition of the word by native listeners" (p. 115). In English, stress only occurs with vowels, not consonants. Stressed vowels in English are characterized by being stronger, louder, longer in duration, greater in intensity, and less centralized in quality compared to non-stressed vowels (Betti & Ulaiw, 2018; Davenport & Hannahs, 2020). English stress comprises varied functions. Davenport and Hannahs (2020) refer to three functions of stress in English: distinction between words, marking contrast, and indicating emphasis. Stress has "a differentiating function"; therefore, the words 'insult', 'record', and 'invalid' have two different readings based on the location of the main stress (p. 80). If the first syllable is stressed in the first two words, we have nouns: 'an **in**sult', 'a **re**cord'; if the second syllable is stressed, we have a verb; 'to in**sult**', 'to re**cord**'. As for the third word, it is either a noun, '**in**valid' (with antepenultimate stress) or an adjective, 'in**va**lid' (with penultimate stress). Stress can be also used to mark contrast, as in 'I said a **big** farm, not a **pig** farm', or to indicate emphasis, as in 'He **walked** to the party yesterday'. The focus of the current study is on the first function of stress, i.e. distinction between words

While both languages are stress-timed, English stress is largely unpredictable due to its mixed Germanic and Romanic origins, whereas Arabic stress is generally predictable (Larudee, 1973; Mitchell, 1960; Watson, 2011; Basalma,

1990). English stress is phonemic, meaning it can change word meanings (Giegerich, 1992; Aziz, 1980), but Arabic stress is non-phonemic (Larudee, 1973; Mitchell, 1960; Sa'di et al., 2022a; Beeston, 2016; Alzi'abi, 2023). These differences often lead Arab learners to transfer Arabic stress patterns to English words, resulting in incorrect stress placement (Anani, 1989; Aziz, 1980; Basalama, 1990; Ghaith, 1993; Helal, 2014; Sa'di et al., 2022a,b). For instance, Arabic speakers tend to stress the penultimate syllable in multisyllabic English words, following Arabic stress rules (Basalama, 1990). The unpredictable nature of English stress and its phonemic value are identified as major sources of difficulty for Arab learners in acquiring English stress patterns (Helal, 2014).

In stress literature, it has been argued that most errors of stress placement in English words by native Arabic speakers are related to the effect of the native language (Anani, 1989; Aziz, 1980; Basalama, 1990; Ghaith, 1993; Helal, 2014; Sa'di et al., 2022a,; Sa'di et al., 2022b). The idea that Arab learners of English have the tendency to transfer stress pattern of their native language into English is asserted by Basalama (1990), who concludes that, on the level of multisyllabic words, such as *Edinburgh, character, severity and Washington*, Arabic speakers shift the stress to the penultimate syllable (second syllable from the end) in conformance with the Arabic stress rules. Aziz (1980), Anani (1989), Ghaith (1993), Helal (2014), Sa'di et al. (2022a), and Sadi et al. (2022b) give various examples of cases of transfer provided by Arabic speakers, as far as English stress is concerned, and all of them agree that native speakers of Arabic carry over their native language fixed word patterns to English test words, consequently, they give incorrect stress placement within English words.

2.6 Learning stress using video captioning

Learning stress using video captioning offers a dynamic approach to comprehending and mastering stress patterns in English. By integrating visual cues with auditory input, learners can better understand the different intonation and rhythm of stress placement within words and sentences. Video captioning provides a multisensory learning experience (Hsieh, 2020), allowing learners to simultaneously see the stress markers while hearing the spoken words. This method enhances learners' ability to recognize and produce correct

stress patterns, crucial for achieving natural and fluent speech. Moreover, the visual representation of stress through captioning aids learners with different learning styles (Chen & Sun, 2012), accommodating diverse needs within language classrooms.

Some studies were conducted to find out the effect of implementing video captioning on learning stress. For example, Maspufah (2022) examined using YouTube videos in teaching word stress pronunciation. The result showed that utilizing YouTube Video improved students' word stress pronunciation. In addition, Boonmee et al., (2022) used closed captions to provide additional visual information on stress placement to help learners make themselves better understood in English conversation. The results showed that the current approach delivered 74.19% of representation accuracy. Also, Sinyashina and Balteiro (2023) used authentic videos with captions to learn the word stress. Ninety Spanish students of English participated in this study. They were asked to watch 5 hours of authentic TV documentary with English captions, where eleven target words had from 2 to 112 repetitions. The results revealed an apparent better performance of the students who viewed captioned authentic videos.

The previous research in video captioning predominantly relied on words to convey the content of the videos. However, phonetic transcription, which represents the sounds of speech, has not been extensively explored as a method for video captioning. This study aims to address this gap in the research by examining and comparing the effectiveness of video captioning using two different approaches: traditional word-based captions versus phonetic transcription-based captions. By considering both categories, this study aims to provide a comprehensive understanding of how different captioning approaches can be leveraged to enhance English language learners' mastery of stress patterns. The findings could inform the development of more effective video captioning strategies, potentially combining elements from both approaches to optimize language learning outcomes. The study will answer the following research question:

RQ: How does the use of phonetic transcription compared to traditional word-based captions in video captioning affect learning word stress?

3. Methods

3.1 Research design

This study employed a quasi-experimental design to examine the impact of captioned videos on English word stress pronunciation among EFL students. The investigation utilized a pre-test and post-test approach to measure the effectiveness of captioned videos on word stress pronunciation.

3.2 Participants

The study sample comprised forty undergraduate EFL students majoring in English Language at the Arab Open University, Saudi Arabia. Participants were randomly assigned to the treatment groups.

3.3 Data collection tools

Pre-test and post-test

The assessment focused on twenty target words (Appendix A), each with two distinct stress patterns depending on their part of speech (e.g., 'desert' as a noun vs. verb). Participants were tasked with reading and recording a set of sentences containing these words using their mobile devices. Each student submitted two audio recordings: one for the pre-test to establish a baseline proficiency level, and another for the post-test to evaluate progress after the intervention. The pre-test was administered before the treatment to assess initial competence, while the post-test followed the training period to measure improvement. These recordings were then sent to instructors for evaluation, providing a comprehensive before-and-after comparison of the students' word stress pronunciation skills.

3.4 Data Evaluation

The instructors evaluated the recordings based on specific criteria, focusing on the accuracy of stress placement in the pronunciation of the targeted words. Each correctly pronounced word was awarded one point, while incorrectly pronounced words received zero points. Thus, the maximum score for each test was twenty points per student. The evaluation aimed to identify any changes or improvements in the students' pronunciation skills.

3.5 Procedure

In this study, pre-test and post-test design was used as key tools to evaluate the impact of different modes of instruction on pronunciation learning. During the first session, students underwent the pre-test by reading and recording a set of twenty short sentences containing the target words (Appendix A). The focus on these specific words in our study was due to several important considerations. First, the selected words exemplify a common and challenging aspect of English pronunciation: the shift in stress depending on the part of speech. For instance, the word "desert" has the stress on the first syllable when used as a noun and on the second syllable when used as a verb. This shift is a fundamental aspect of English pronunciation that often poses difficulties for learners, making it a suitable focus for evaluating the impact of instructional methods. Second, while it is true that there are many pronunciation rules in English, a focused approach allows for a more controlled and detailed investigation. By concentrating on a specific subset of words with varying stress patterns, we can more accurately measure the effects of the instructional interventions. Broadening the scope to include a wider range of pronunciation rules might dilute the focus and make it more challenging to draw precise conclusions. Lastly, the choice of these words was informed by their relevance and frequency in everyday language use, ensuring that the findings of our study are practical and applicable to real-world language learning contexts.

This pre-test served as a baseline measure of pronunciation proficiency before any instructional intervention took place. In the second session, participants were divided into two groups: an experimental group and a control group. Both groups were exposed to the same content, but with different modes of instruction. In the experimental group, participants viewed a video containing sentences with target words, with these words captioned in phonetic transcription (See **Figure 1**). This mode aimed to provide learners with a visual representation of the sounds of the words. On the other hand, the control group viewed a similar content, but the target words were captioned using traditional spelling (See **Figure 2**). This mode served as a comparison to the experimental group, focusing on the visual representation of the words' conventional orthography. At the end of the second session, participants in both groups underwent the post-test, where they were asked to read aloud the same set of twenty

sentences containing the target words. This post-test served to measure any improvement in pronunciation proficiency after exposure to the different modes of instruction.
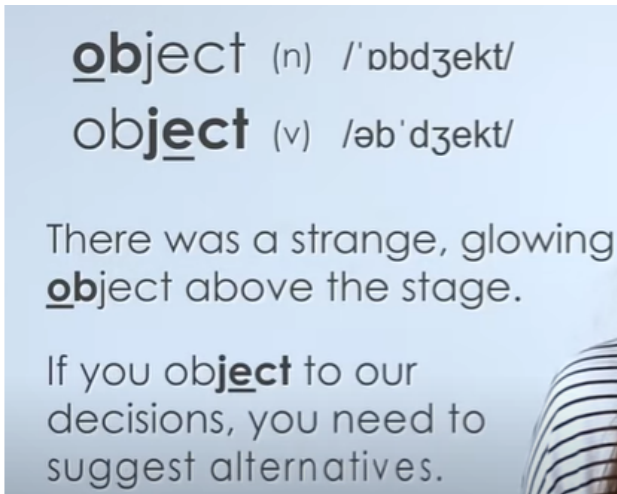

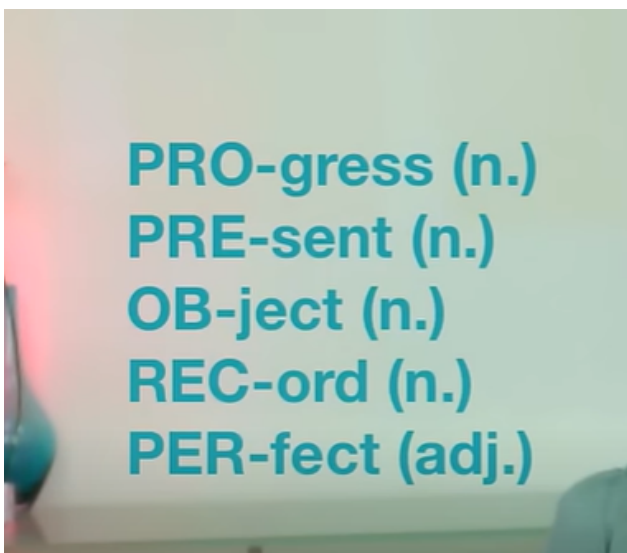**Figure 1.** A screenshot of the video used for the experimental group.

Figure 2. A screenshot of the video used for the control group.

3.6 Data Analysis

First, a test of normality to assess whether a dataset follows a normal distribution was conducted. This test helps researchers determine the appropriateness of parametric statistical methods for their analyses. In this study, the data was not normally distributed as shown in **Table 1**. The Kolmogorov-Smirnov test and the Shapiro-Wilk test showed that all values were less than 0.05. This indicated that the data was normally distributed. Thus, non-parametric tests

should be used to analyze the data.

To comprehensively address the research question posed in this study, rigorous analysis was conducted of the students' pronunciation test scores both before and after the intervention. To address the first research question, we meticulously examined the data using a non-parametric test for comparing the means of two groups. Independent-Samples Mann-Whitney U Test was used to find out the difference between the scores of the pre-test and post-test of both groups. The Independent-Samples Mann-Whitney U Test is a non-parametric statistical test used to compare the means of two independent groups. It is particularly useful when the assumptions of the independent samples t-test are not met or when dealing with ordinal or non-normally distributed data. This statistical approach enabled a comparison of scores obtained by participants in the two distinct groups before and after the intervention, aiming to uncover any significant differences in pronunciation. All these analyses were conducted using SPSS version 29 (IBM Corp., 2020).

4. Results

4.1 Descriptive analysis

The results shown in **Table 2** present the mean, standard deviation, and standard error of the mean for pre-test and post-test scores of the two groups: control and experimental. In the pre-test, the control group had a mean score of 12.20 with a standard deviation of 6.65, while the experimental group had a slightly higher mean of 12.45 with a standard deviation of 5.10. In the post-test, both groups showed an improvement in scores, with the control group having a mean of 16.80 and the experimental group having a mean of 16.35. The standard deviations for the post-test scores are lower, indicating less variability. The standard error of the mean for both pre-test and post-test scores reflects the precision of the sample mean estimates. The results suggested that both groups experienced an increase in scores from the pre-test to the post-test, with similar means.

The Mann-Whitney U test was used to compare two independent groups, and the results indicated that there was no statistically significant difference between the groups at the specified significance level. Therefore, based on these findings, there was insufficient evidence to conclude that the groups differ on the measured variable.

Table 1. Test of normality.

	Group	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Pre-test	Control	0.215	20	0.016	0.849	20	0.005
	Experiment	0.241	20	0.003	0.856	20	0.007
Post-test	Control	0.196	20	0.042	0.851	20	0.006
	Experiment	0.266	20	0.001	0.863	20	0.009

Table 2. Descriptive analysis.

	Group	N	Mean	Std. Deviation	Std. Error Mean
Pre-test	Control	20	12.2000	6.64593	1.48608
	Experiment	20	12.4500	5.10392	1.14127
Post-test	Control	20	16.8000	3.31821	0.74197
	Experiment	20	16.3500	2.20705	0.49351

Table 3 presents the results of an Independent-Samples Mann-Whitney U Test comparing the pre-test scores of both groups. With a total sample size of forty participants, the test yielded a Mann-Whitney U statistic of 195.500 and a Wilcoxon W of 405.500. The standardized test statistic is −0.122, which is very close to zero. Most importantly, the asymptotic significance (2-sided test) is 0.903, and the exact significance (2-sided test) is 0.904. Both of these p-values are well above the conventional significance level of 0.05. These results suggest that there was no statistically significant difference between the two groups in their pre-test scores. The high p-values (0.903 and 0.904) indicate that any observed differences between the groups at the pre-test stage were likely due to chance rather than any systematic difference between the groups. This supports the assumption that the groups were equivalent at the beginning of the study, which is crucial for interpreting any post-test differences.

4.2 Post-test

Table 4 shows the results of the Independent-Samples Mann-Whitney U Test for the post-test scores of both groups. Again, the total sample size is forty. The Mann-Whitney U statistic is 168.500, and the Wilcoxon W is 378.500. The standardized test statistic is -0.859, which indicates a slightly larger difference than in the pre-test, but still not substantial. The asymptotic significance (2-sided test) is 0.390, and the exact significance (2-sided test) is 0.398. These results indicate that there was no statistically significant difference between the two groups in their post-test scores. The p-values (0.390 and 0.398) are lower than those in the pre-test, suggesting a slightly larger difference between the groups, but they are still well above the 0.05 significance level.

The results from both tables suggest that there was no statistically significant difference between the two groups at either the pre-test or post-test stage. This implies that the intervention (presumably the different types of video captioning) did not lead to a significant difference in performance between the groups. The lack of significant difference between groups does not necessarily mean the intervention was ineffective; it might mean that both interventions were equally effective. A Cohen"s d value of 0.160 suggested a small effect size. This indicated that there was a measurable difference between the groups, but it was relatively small in magnitude. It implied that while there might be a statistically significant difference between the groups, it may not have a substantial practical or clinical significance.Top of Form

5. Discussion

The primary aim of this study was to investigate how video captioning, employing traditional word-based captions versus phonetic transcription-based captions, impacts EFL learners" pronunciation of English word stress. The findings reveal that both the control group, exposed to captions featuring spelled words; and the experimental group, exposed to captions with phonetic transcription, demonstrated notable enhancements in their English stress pronunciation scores

Table 3. Independent-Samples Mann-Whitney U Test Summary of pre-test of both groups.

Total N	40
Mann-Whitney U	195.500
Wilcoxon W	405.500
Test Statistic	195.500
Standard Error	36.798
Standardized Test Statistic	−0.122
Asymptotic Sig. (2-sided test)	0.903
Exact Sig. (2-sided test)	0.904

N = number of the participants.

Table 4. Independent-Samples Mann-Whitney U Test Summary of post-test of both groups.

Total N	40
Mann-Whitney U	168.500
Wilcoxon W	378.500
Test Statistic	168.500
Standard Error	36.655
Standardized Test Statistic	−0.859
Asymptotic Sig. (2-sided test)	0.390
Exact Sig. (2-sided test)	0.398

N = number of the participants.

following the post-test assessment. This improvement across both groups highlights the usefulness and efficacy of video captioning as a compelling tool for supporting pronunciation skills among EFL learners. The presence of written text alongside the spoken content within the video seems to have fostered progress for all participants, regardless of the specific captioning format utilized. This finding aligns with contemporary research emphasizing the multimodal nature of language acquisition and the benefits derived from integrating visual and auditory stimuli.

Our results are consistent with several studies in the literature. For instance, Wisniewska and Mora (2020) found that extended exposure to captioned videos benefited speech segmentation and speech processing skills, regardless of the viewing mode. Similarly, Mahdi (2017) demonstrated that both keyword captioning and full captioning were effective in enhancing L2 pronunciation. These studies, along with our findings, emphasize the crucial role of video captioning in facilitating effective language learning strategies and highlight its potential to enhance pronunciation skills in diverse educational settings.

The improvement observed in both groups can be theoretically supported by the Dual Coding Theory proposed by Paivio (2014). This theory suggests that humans process and retain information more effectively when it is presented through multiple sensory channels, particularly visual and auditory. In our study, the video captions served as a powerful tool for language learners, providing them with simultaneous visual reinforcement alongside auditory input. This dual coding process likely facilitated deeper encoding of linguistic information, enhancing learners' ability to recall and reproduce pronunciation accurately.

Interestingly, while both groups improved, the learners exposed to video captioned in spelled words scored slightly higher in the post-test compared to those exposed to phonetic transcription. This subtle difference might be attributed to learners' greater familiarity with traditional spelling compared to phonetic symbols, potentially reducing cognitive load and allowing them to focus more on pronunciation itself (Nation, 2003). However, it is important to note that this difference was relatively small and may not be statistically significant.

The descriptive statistics, coupled with the Mann-Whitney U test results for the pretest scores, justify the assumption of homogeneity at baseline. This homogeneity is crucial for attributing any post-test differences to the intervention rather than pre-existing disparities. The increase in mean scores from the pretest to the post-test for both groups

suggests that some form of improvement occurred. However, since both groups improved similarly and the Mann-Whitney U test for the post-test scores shows no significant difference, it implies that the intervention did not have a differential effect between the groups. These findings align with Mohsen and Mahdi (2021), who found that both full and partial captioning groups outperformed the non-captioning group in pronunciation tests, with no significant difference between the two captioning types. This suggests that video captioning, regardless of its specific format, can improve word pronunciation in second language acquisition.

Our study contributes to the growing body of research on the use of technology in language learning, particularly in the realm of pronunciation improvement. It supports the notion that extended exposure to captioned videos can be beneficial for second language pronunciation, as long as learners' attention is appropriately directed (Wisniewska & Mora, 2020).

Future research could explore the long-term effects of different captioning types on pronunciation skills and investigate potential interactions with learners' proficiency levels and cognitive styles. Additionally, examining the impact of video captioning on other suprasegmental features of pronunciation, such as intonation and rhythm, could provide a more comprehensive understanding of its benefits for language learners.

6. Conclusion

This study investigated the impact of video captioning, utilizing traditional word-based captions versus phonetic transcription-based captions, on EFL learners' pronunciation of English stress. The results demonstrated notable enhancements in English stress pronunciation scores for both the control group, exposed to spelled words, and the experimental group, exposed to phonetic transcription, post-test assessment. This underscores the effectiveness of video captioning as a valuable tool for supporting pronunciation skills among EFL learners, irrespective of the captioning format utilized. However, it is worth noting that learners exposed to video captioned with spelled words scored slightly higher in the post-test compared to those exposed to phonetic transcription. This difference may stem from learners' familiarity with traditional spelling versus decoding phonetic symbols,

potentially reducing cognitive load and allowing for greater focus on pronunciation. Further research with larger sample sizes and robust statistical analyses is warranted to validate these findings definitively.

The findings underscore the pedagogical value of incorporating video captioning into language instruction to enhance pronunciation skills among EFL learners. Educators can integrate video captioning activities into lesson plans to provide learners with additional visual reinforcement alongside auditory input, facilitating deeper comprehension and memory retention of pronunciation patterns. Pedagogical practices should embrace multimodal approaches that leverage the integration of visual and auditory stimuli in language learning materials. By integrating text with spoken language, educators can capitalize on the principles of Dual Coding Theory to optimize pronunciation learning outcomes and enhance learners' ability to recall and reproduce pronunciation accurately. In addition, educators should consider tailoring captioning formats to accommodate learners' needs and preferences. While both traditional word-based captions and phonetic transcription-based captions yield improvements in pronunciation, instructors may opt to provide learners with a variety of captioning formats to cater to diverse learning styles and linguistic backgrounds. Also, language instruction should emphasize the development of transferable skills that enable learners to connect spelling knowledge with pronunciation patterns. Educators can design activities that encourage learners to make explicit connections between written representations of words and their corresponding sounds, fostering a deeper understanding of pronunciation and promoting transferable skills that enhance overall language proficiency.

The study has some limitations. First, the sample size is small. The study may not adequately represent the diverse range of EFL learners, potentially limiting the generalizability of the findings to broader populations of language learners. Second, the study's duration of exposure to video captioning may not have been sufficient to capture long-term effects on pronunciation learning. Longer-term studies could provide insights into the sustainability of pronunciation improvements over time and the optimal duration of exposure needed for significant gains. Third, the post-test assessment solely focusing on English stress pronunciation may not fully capture the comprehensive impact of video captioning on overall pronunciation proficiency. Including a

broader range of pronunciation measures, such as intonation or segmental accuracy, could provide a more comprehensive understanding of the effects of video captioning on pronunciation learning. Fourth, the study did not include a delayed post-test to measure the effect of video captioning on the retention of word stress among the participants. The study employed two groups utilizing video captioning as an intervention. A more robust design would involve three groups, including a control group that did not use any form of video captioning. Additionally, the experiment's scope was constrained by the limited number of words tested, specifically twenty words. Future studies can be conducted with three groups and more words to fulfill the generalizability of the findings.

Therefore, future studies are required to address these limitations. For example, future studies could be conducted with a larger sample size and composition by recruiting a larger and more diverse sample of EFL learners. Researchers could include participants from various age groups, language backgrounds, proficiency levels, and learning contexts to ensure the findings are more representative and applicable to a broader range of language learners. Also, to address concerns about the duration of exposure, researchers could conduct longitudinal studies with extended periods of video captioning exposure and with applying delayed post-tests. By tracking pronunciation improvements over an extended timeframe, researchers can assess the sustainability of gains and determine the optimal duration of exposure needed for significant and lasting improvements in pronunciation. Moreover, future studies could expand the scope of assessment methods beyond English stress pronunciation to include a broader range of pronunciation measures. This could involve assessing intonation patterns, rhythm, and other aspects of pronunciation to provide a more comprehensive understanding of the effects of video captioning on overall pronunciation proficiency.

Author Contributions

SMMH wrote original draft, HSM did the analysis, MAT revised the paper. All authors have read and agreed to the published version of manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study will be available upon reasonable request.

Funding

Acknowledgments

Appendix A. Pronunciation test

Read the following sentences. Pay attention to the words written in bold.

1. A. They were in **conflict**.
   B. Your views **conflict** with mine.
2. A. The **contrast** between them is high.
   B. I will **contrast** the two offers.
3. A. Coffee is an **import** from Brazil.
   B. We will **import** more coffee.
4. A. Do you have a **permit** to go?
   B. Will you **permit** me to leave?
5. A. She gave me a **present**.
   B. Allow me to **present** my friend.
6. A. I have a **record** of the class.
   B. Could you please **record** the class?
7. A. Do you know his **address**?
   B. I'll **address** three questions.
8. A. We are happy with your **conduct**.
   B. Please, **conduct** yourself well.
9. A. They drove through the **desert**.
   B. His family may **desert** him.
10. A. He became an **invalid** in the last war.
    B. The accident might **invalid** her.

References

Abbas, N., Ali Qassim, T., 2020. Investigating the effectiveness of YouTube as a learning tool among EFL students at Baghdad University. Arab World En-

glish Journal (AWEJ) Special Issue on CALL. (6).

Abdar, M., Kollati, M., Kuraparthi, S., et al., 2023. A review of deep learning for video captioning. arXiv preprint arXiv:2304.11431.

Alkhuli, M.A., 2002. A contrastive transformational grammar. Al Manhal.

Alwehaibi, H.O., 2015. The Impact of Using Youtube in EFL Classroom on Enhancing EFL Students' Content Learning. Journal of College Teaching & Learning. 12(2), 121–126.

Alzi'abi, S.E., 2023. Arab learners' stress perception and production of English multisyllable items. Journal of Experimental Phonetics. 32, 111–130.

Anani, M., 1989. Incorrect Stress Placement in the Case of Arab Learners of English. IRAL: International Review of Applied Linguistics in Language Teaching. 27(1), 15–22. DOI: https://doi.org/10.1515/iral.1989.27.1.15.

Avello, D., Muñoz, C., 2023. The development of receptive language skills from captioned video viewing in primary school EFL learners. Education Sciences. 13(5), 479. DOI: https://doi.org/10.3390/educ- sci13050479.

Aziz, Y.Y., 1980. Some Problems of English Word-Stress for the Iraqi Learner. English Language Teaching Journal. 34(2), 104–109. DOI: https://doi.org/10.1093/elt/34.2.104.

Basalama, M.S., 1990. A study of certain aspects of suprasegmental of Arabic, and their influence on the pronunciation of Arab teachers from the Western Zone of Saudi Arabic [PhD thesis]. London: University of London.

Beeston, A.F.L., 2016. The Arabic language today. Routledge.

Betti, M.J., Ulaiwi, W.A., 2018. Stress in English and Arabic: A contrastive study. English Language and Literature Studies. 8(1), 83–91. DOI: https://doi.org/10.5539/ells.v8n1p83.

Bird, S.A., Williams, J.N., 2002. The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. Applied psycholinguistics. 23(4), 509–533. DOI: https://doi.org/10.1017/S0142716402004022.

Boonmee, P., Jitngernmadan, P., Kubola, K., et al., 2022. Accuracy of Natural Language Processing Algorithms in Represent- ing Stressed Words in Closed Captioning. In 2022 27th Asia Pacific Conference on Commu- nications (APCC) (pp. 348–352). IEEE. DOI: https://doi.org/10.1109/APCC55198.2022.9943727.

Borrás, I., Lafayette, R.C., 1994. Effects of multimedia courseware subtitling on the speaking performance of college students of French. The Modern Language Journal. 78(1), 61–75. DOI: https://doi.org/10.2307/329253.

Cengiz, B.C., 2023. Computer-assisted pronunciation teaching: An analysis of empirical research. Participatory Educational Research. 10(3), 72–88. DOI: https://doi.org/10.17275/per.23.45.10.3.

Chai, J., Erlam, R., 2008. The effect and the influence of the use of video and captions on second language learning. New Zealand Studies in Applied Linguistics. 14(2), 25–44.

Chen, C.M., Sun, Y.C., 2012. Assessing the effects of different multimedia materials on emotions and learning performance for visual and verbal style learners. Computers & Education. 59(4), 1273–1285. DOI: https://doi.org/10.1016/j.compedu.2012.05.006.

Chun, D.M., Plass, J.L., 1997. Research on text comprehension in multimedia environments. Language Learning & Technology. 1(1), 60–81.

Chung, J.M., 1999. The effects of using video texts supported with advance organizers and captions on Chinese college students' listening comprehension: An empirical study. Foreign Language Annals. 32(3), 295–308. DOI: https://doi.org/10.1111/j.1944-9720.1999.tb01342.x.

Cutler, A., 2015. Lexical stress in English pronunciation. The handbook of English pronunciation. 106–124. DOI: https://doi.org/10.1002/9781118346952.ch6.

Danan, M., 1992. Reversed subtitling and dual coding theory: New directions for foreign language instruction. Language learning. 42(4), 497–527. DOI: https://doi.org/10.1111/j.1467-1770.1992.tb01042.x.

Danan, M., 2004. Captioning and subtitling: Undervalued language learning strategies. Meta. 49(1), 67–77. DOI: https://doi.org/10.7202/009021ar.

Dao, P., Nguyen, M.X.N.C., Nguyen, N.B.C., 2021. Effect of pronunciation instruction on L2 learners' listening comprehension. Journal of Second Language Pronunciation. 7(1), 10–37. DOI: https://doi.org/10.1075/jslp.19012.dao.

Darcy, I., Rocca, B., 2022. Comprehensibility improvements in integrated pronunciation instruction: A comparison of instructional methods and task effects. Journal of Second Language Pronunciation. 8(3), 328–362. DOI: https://doi.org/10.1075/jslp.21035.dar.

Davenport, M., Hannahs, S.J., 2020. Introducing phonetics and phonology (4th ed.). Routledge. DOI: https://doi.org/10.4324/9781351042789.

Davey, R., Parkhill, F., 2012. Raising adolescent reading achievement: The use of sub-titled popular movies and high interest literacy activities. English in Aotearoa. (78), 61–71.

Duffy, P., 2008. Using Youtube: Strategies for using new media in teaching and learning. In Enhancing learning through technology: research on emerging technologies and pedagogies (pp. 31–43). DOI: https://doi.org/10.1142/9789812799456_0003.

Garza, T.J., 1991. Evaluating the use of captioned video materials in advanced foreign language learning. Foreign language annals. 24(3), 239–258. DOI: https://doi.org/10.1111/j.19449720.1991.tb00469.x.

Gernsbacher, M.A., 2015. Video captions benefit everyone. Policy insights from the be- havioral and brain sciences. 2(1), 195–202. DOI: https://doi.org/10.1177/2372732215602130.

Ghaith, S., 1993. The assignment of primary stress to words by some Arab speakers. System. 21(3), 381–390. DOI: https://doi.org/10.1016/0346251X(93)90028-F.

Ghosh, M., Levis, J.M., 2021. Vowel quality and direction of stress shift in a predictive model explaining the varying impact of misplaced word stress: Evidence from English. Frontiers in Communication. 6, 628780. DOI: https://doi.org/10.3389/fcomm.2021.628780.

Giegerich, H.J., 1992. English phonology: An introduction. Cambridge: Cambridge University Press.

Helal, S., 2014. Stress in English: Prosodic and rhythmic complexity for Arab learners. Concordia Working Papers in Applied Linguistics. 5, 361–294.

Hogg, R.M., McCully, C.B., 1987. Metrical phonology: a course book. Cambridge University Press.

Hsieh, Y., 2020. Effects of video captioning on EFL vocabulary learning and listening comprehension. Computer Assisted Language Learning. 33(5–6), 567–589. DOI: https://doi.org/10.1080/09588221.2019.1577898.

Hsu, C.K., Hwang, G.J., Chang, Y.T., et al., 2013. Effects of video caption modes on English listening comprehension and vocabulary acquisition using hand-held devices. Journal of Educational Technology & Society. 16(1), 403–414.

Huwari, I.F., Mehawesh, M., 2015. Review of the importance of teaching pronunciation in the Arab society. International Journal on Studies in English Language and Literature (IJSELL). 3(6), 31–37.

Hwang, Y.L., 2003. The effect of the use of videos captioning on English as a foreign language (EFL) on college students' language learning in Taiwan. The University of Mississippi.

IBM Corp., 2020. IBM SPSS Statistics for Windows (Version 29.0) [Computer software]. IBM Corp. https://www.ibm.com/analytics/spss-statistics-software.

Jenkins, J., 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. Applied linguistics. 23(1), 83–103. DOI: https://doi.org/10.1093/applin/23.1.83.

Kreidler, C.W., 2008. The pronunciation of English: A course book. John Wiley & Sons.

Ladefoged, P., Johnson, K., 2014. A Course in Phonetics. Belmont, CA: Cengage Learning.

Ladefoged, P., 2011. A course in phonetics (6th ed.). Cengage Learning.

Larudee, F., 1973. Word Stress in the Spoken Arabic of Cairo. Language Sciences. 26, 31–34.

Lo, S., 2024. Vocabulary learning through viewing dual-subtitled videos: Immediate repetition versus spaced repetition as an enhancement strategy. ReCALL. 1–16.

Mahdi, H.S., 2017. The use of keyword video captioning on vocabulary learning through mobile-assisted language learning. International journal of English linguistics. 7(4), 1–7. DOI: https://doi.org/10.5539/ijel.v7n4p1.

Markham, P., 1999. Captioned videotapes and second-language listening word recognition. Foreign Language Annals. 32(3), 321–328. DOI: https://doi.org/10.1111/j.1944-9720.1999.tb01344.x.

Markham, P., 2001. The influence of culture-specific background knowledge and captions on second language comprehension. Journal of Educational Technology Systems. 29(4), 331–343. DOI: https://doi.org/10.2190/15TA-GX8P-74XP-YUA1.

Markham, P.L., Peter, L.A., McCarthy, T.J., 2001. The effects of native language vs. tar- get language captions on foreign language students' DVD video comprehension. Foreign language annals. 34(5), 439–445. DOI: https://doi.org/10.1111/j.19449720.2001.tb-02083.x.

Markham, P., Peter, L., 2003. The influence of English language and Spanish language captions on foreign language listening/reading comprehension. Journal of Educational Technology Systems. 31(3), 331–341. DOI: https://doi.org/10.2190/BHUH-420B-ALA0.

Maspufah, M., 2022. Utilizing YouTube video as media in teaching students' word stress pronunciation. Pendekar: Jurnal Pendidikan Berkarakter. 5(4), 23–28.

Mayer, R.E., 2003. The promise of multimedia learning: using the same instructional design methods across different media. Learning and instruction. 13(2), 125–139. DOI: https://doi.org/10.1016/S0959(02)000-16-6.

Mitchell, T.F., 1960. Prominence and Syllabication in Arabic. Bulletin of the School of Oriental and African Studies. 23(2), 369–389. DOI: https://doi.org/10.1017/S0041977X00149997.

Mitterer, H., McQueen, J.M., 2009. Foreign subtitles help but native-language subtitles harm foreign speech perception. PloS one. 4(11), e7785. DOI:

https://doi.org/10.1371/journal.pone.0007785.

Mohsen, M.A., Mahdi, H.S., 2021. Partial versus full captioning mode to improve L2 vocabulary acquisition in a mobile-assisted language learning setting: words pronunciation domain. Journal of Computing in Higher Education. 33(2), 524–543. DOI: https://doi.org/10.1007/s12528-021-09276-0.

Morley, J., 1991. The pronunciation component in teaching English to speakers of other languages. TESOL quarterly. 25(3), 481–520. DOI: https://doi.org/10.2307/3586981.

Nation, I.S.P., 2001. Learning vocabulary in another language (Vol. 10, pp. 126–132). Cambridge: Cambridge University Press.

Paivio, A., 2014. Mind and its evolution: A dual coding theoretical approach. Psychology press.

Perez, M.M., 2022. Second or foreign language learning through watching audiovisual input and the role of on-screen text. Language Teaching. 55(2), 163–192. DOI: https://doi.org/10.1017/S0261444821000501.

Perez, M.M., Peters, E., Desmet, P., 2014. Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. ReCALL. 26(1), 21–43. DOI: https://doi.org/10.1017/S0958344013000256.

Plass, J.L., Chun, D.M., Mayer, R.E., et al., 1998. Supporting visual and verbal learning preferences in a second-language multimedia learning environment. Journal of educational psychology. 90(1), 25.

Qiao, W., Yijun, C., 2023. The role of live transcripts in synchronous online L2 classrooms: Learning outcomes and learner perceptions. Education and Information Technologies. 28(11), 14783–14804. DOI: https://doi.org/10.1007/s10639-023-11784-8.

Roach, P., 2009. English phonetics and phonology paperback with audio CDs (2): A practical course. Cambridge University Press.

Sa'di, R.A., Sharadgah, T.A., Yaseen, M.S., 2022a. Stress Misassignment in the Pronunciation of English by Arabic-speaking Learners: Erratic Practice or Crosslinguistic Influence? International Journal of Arabic-English Studies (IJAES). 22(1). DOI: https://doi.org/10.33806/ijaes2000.22.1.5.

Sa'di, R.A., Sharadgah, T.A., Yaseen, M.S., 2022b. An Investigation of Common Stress Placement Errors in English Word Roots and their Suffixed Derivatives by Arabic-Speaking EFL Learners. Arab World English Journal. 13(3). DOI: https://doi.org/10.24093/awej/vol13no3.4.

Sharma, L.R., 2023. Dealing with the word and sentence stress in English language. International Journal of Multidisciplinary Comprehensive Research. 2(1), 31–37. DOI:

https://doi.org/10.54660/ijmcr.2023.2.1.31-37.

Sinyashina, E., Balteiro, I., 2023. Incidental Learning of Word Stress with Captioned Authentic Videos: The Effect of Repetition. RAEL: Revista Electrónica de Lingüística Aplicada. 21(1), 19–39. DOI: https://doi.org/10.58859/rael.v21i1.494.

Steinfeld, A., 1998. The benefit of real-time captioning in a mainstream classroom as measured by working memory. Volta review. 100(1).

Syodorenko, T., 2010. Modality of input and vocabulary acquisition. Language Learning & Technology. 14, 50–73.

Teng, M.F., 2023. Incidental vocabulary learning from captioned videos: Learners' prior vocabulary knowledge and working memory. Journal of Computer Assisted Learning. 39(2), 517–531. DOI: https://doi.org/10.1111/jcal.12756.

Tohe, A., 2021. YouTube, Learning, and Transformative Critical Pedagogy. KnE Social Sciences. 15–29. DOI: https://doi.org/10.1017/S0272263120000029.

Treiman, R., Kessler, B., 2023. Spelling-to-sound translation for English disyllables: Use of long and short vowels before single medial consonants. Journal of Experimental Psychology: Learning, Memory, and Cognition. 49(12), 2034–2047. DOI: https://doi.org/10.1037/xlm0001260.

Vázquez, L.Q., 2022. Synchronous computer-mediated communication in pronunciation teaching: A survey study. Revista de Estilos de Aprendizaje. 15, 123–134. DOI: https://doi.org/10.55777/rea.v15iEspecial.4431.

Watson, J.C.E., 2011. Word stress in Arabic. In M. Oostendorp, C. Ewen, E. Hume, et al. (eds.). The Blackwell companion to phonology (Vol. 5) (pp. 2990–3019). Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444335262.wbctp0124.

Winke, P., Gass, S., Sydorenko, T., 2010. The effects of captioning videos used for foreign language listening activities. Language Learning & Technology. 14, 65–86.

Wisniewska, N., Mora, J.C., 2017. Pronunciation learning through captioned videos. Pronunciation in Second Language Learning and Teaching Proceedings. 9(1).

Wisniewska, N., Mora, J.C., 2020. Can captioned video benefit second language pronunciation? Studies in Second Language Acquisition. 42(3), 599–624. DOI: https://doi.org/10.1017/S0272263120000029.

Yeldham, M., 2018. Viewing L2 captioned videos: what's in it for the listener?. Computer Assisted Language Learning. 31(4), 367–389. DOI: https://doi.org/10.1080/09588221.2017.1406956.