



<https://jrl.ui.ac.ir/?lang=en>

Journal of Researches in Linguistics
E-ISSN: 2322-3413
16(2), 41-52
Received: 26.04.2024 Accepted: 20.07.2024

Research Paper

A virtual-reality (VR) cognitive pupillometry analysis of auditory and visual phonemic awareness tasks involving ‘th’ sound variations

Mohsen Mahmoudi-Dehaki*

English Department, Najafabad Branch, Islamic Azad University, Najafabad, Iran
mohsen.mahmoudi@shu.iaun.ac.ir

Nasim Nasr-Esfahani

Department of English, Isfahan (khorasgan) Branch, Islamic Azad University, Isfahan, Iran
nasresfahanitefl@gmail.com

Abstract

Rooted in the Cognitive Load (CL) and Dual Coding theories, this sequential explanatory study examined the CL levels experienced by 36 available Iranian pre-intermediate English as a Foreign Language (EFL) learners in auditory and visual Phonemic Awareness Tasks (PATs). The PATs focused on 'th' sound variations that were specifically challenging for Persian speakers of English. Participants were divided into two equal groups. The first group underwent auditory and visual PATs (i.e., picture sorting and word discrimination). The second group faced auditory PATs only (i.e., rhyme recognition and sound isolation) structured around voiced and voiceless 'th' sounds. In the qualitative phase, 8 participants were randomly interviewed to discover the potential reasons behind various levels of CL measured in certain PATs. A Virtual Reality (VR) cognitive pupillometry analysis was performed for two groups in an immersive VR environment using VR Head-Mounted Displays (VR-HMDs) and Tobii Nexus software, followed by thematic analysis. GazeMetrics strategies were employed to validate and ensure the reliability of the data collection process. The quantitative results demonstrated (a) no critical differences regarding the CL experienced by the participants in each group concerning voiced vs. voiceless 'th' PATs and (b) higher levels of CL (increased pupil diameter, decreased blink rates, and increased gaze pattern or fixation duration) among participants in the auditory PATs group compared to the ones in the auditory and visual PATs group. Besides, qualitative findings identified the following factors as the most probable causes of the observed differences in CL levels among participants: (a) segmental differences, (b) phonological transfer, (c) phonological fossilization, (d) working memory constraints, (e) cognitive flexibility, and (f) task delivery methods. This study contributes to cognitive phonology, phonological processing, perceptual-motor language functions, and English material development.

Keywords: Cognitive load; Cognitive pupillometry; Phonemic awareness task; Phonological processing; Virtual reality

1. Introduction

Phonemic Awareness (PA) is an ability to identify, isolate, and manipulate individual sounds or phonemes in spoken words. Phonemic awareness is critical to learning English since it plays a fundamental role in assisting learners to understand the relationship between letters and sounds (Alswede, 2022). Extensive research in this connection (e.g., Djiguimkoudre, 2021; Fabre-Merchan et al., 2017) have emphasized the importance of meticulous studies on the construction and manipulation of Phonemic Awareness Tasks (PATs). The primary objective of such tasks is to assess phonemic awareness (PA) validly and reliably and yield valuable insights from a pedagogical standpoint and various cognitive domains. This is particularly salient in the context of English as a Foreign Language (EFL) education, where the presence of phonemic distinctions, not being present in the learner's first language (L1), can create significant

*Corresponding author



challenges. Therefore, PATs must be designed and implemented to account for these complexities and facilitate acquiring PA skills in EFL contexts.

According to Dual Route Cascaded (DRC) model, a cognitive model of reading, there are two pathways for reading: the phonological route (sounding out words) and the lexical route (recognizing words by sight). PA primarily supports the phonological route by enabling learners to segment and blend sounds, facilitating early reading development and spelling (Coltheart et al., 2001). Thus, PA is crucial for decoding and translating written text into spoken words. Besides, cognitive scientists argue that PA is crucial in vocabulary development. By breaking down words into their constituent sounds, learners can more easily access and store them in their mental lexicon. This process aids the learners not only in word recognition but also in word learning, as it allows individuals to infer meanings based on phonemic similarities between known and unknown words (Ehri et al., 2001). Moreover, research within cognitive science has consistently found PA to be a strong predictor of reading success because it is one of the critical components for effective reading instruction (Castles & Coltheart, 2004). EFL learners who develop strong PA skills tend to become proficient readers and spellers, as understanding the phonemic structure of words is essential for mapping sounds onto print.

Furthermore, English has a deep orthography. This means that there is not always a straightforward correspondence between letters and sounds. PA helps learners navigate this complexity by providing the skills to understand and apply these relationships (Rokhman, 2020). Cognitive theories such as connectionism suggest that exposure to language patterns, including phoneme-grapheme correspondences, strengthens neural connections that support language processing. Finally, while PA directly influences decoding skills, it indirectly supports reading comprehension (Griffith & Olson, 1992). Cognitive models, like the Simple View of Reading, propose that decoding and linguistic comprehension are necessary for reading comprehension. Through facilitating efficient decoding, PA allows readers to allocate more cognitive resources to comprehension processes (Tan et al., 2007). Apart from the essential advantages of enhancing PA among EFL learners, the valid and reliable approach in assessing EFL learners' Cognitive Load (CL) is the mental effort required to process novel information and integrate it with existing knowledge. This is while performing PATs provide significant insights in regard with the cognitive processes in English learning.

Managing CL prevents overloading of working memory, the part of short-term memory concerned with immediate conscious perceptual and linguistic processing (Kanokpermpoon, 2013). This efficiency is crucial for language learning since it allows for better retention and recall of new vocabularies and grammar rules. In addition, provided that the learners are not being overwhelmed by CL, they can more effectively transfer their knowledge about phonemes to new words and contexts (Smalle et al., 2021). Such an ability to generalize learning is a critical component of language acquisition. Ultimately, measuring the CL as well as being adjusted for it contributes learners to developing metacognitive strategies to manage their learning process (Burns, 2023). Learners also become more aware of their learning requirements and how best to address them, which is beneficial for lifelong learning. Considering the Iranian EFL context, Persian English speakers often encounter challenges in regard with distinguishing between voiced and unvoiced 'th' sounds due to differences in phonetic systems. Therefore, it is essential to understand the CL levels involved in different types of PATs through Virtual Reality (VR) cognitive pupillometry analysis. This is because it plays a pivotal role in designing effective language learning interventions from various scientific disciplines, primarily cognitive linguistics and cognitive psychology. What's more, it examines pupil dilation, blink rates, and gaze patterns as indicators of CL, which offers objective, real-time data on cognitive effort and provides a more nuanced understanding of how EFL learners process challenging phonemic contrasts.

2. Research Background

2.1 Theoretical Background

From a theoretical perspective, this study is fundamentally supported by various dominant theories such as Cognitive Load Theory (CLT), Dual Coding Theory (DCT), and Pupillary Light Reflex Theory (PLRT). CLT posits that our working memory is limited to processing new information. CL refers to the complexity inherent in the material learnt independent of the instructional design. When learning to identify phonemic distinctions while completing PATs, the cognitive load may be influenced by the complexity of the specific sounds being differentiated. For instance, distinguishing between the voiced 'th' sound in words like "this" and the unvoiced 'th' sound in words like "think" could increase the cognitive demands compared to identifying other types of phonemic contrasts. DCT suggests that auditory and visual information are processed in two distinct channels in human mind. Each of these channels is endowed with particular capacity limits. When applied to auditory and auditory/visual tasks for PA, DCT supports the idea that presenting information both through auditory and visual ways (e.g., showing a picture of a thumb while saying "thumb") can reduce the CL by distributing processing across both channels, making it easier to learn and distinguish between different 'th' sounds.

Cognitive pupillometry analysis is employed in cognitive science and psychology to measure the pupil's diameter as an indicator of cognitive activity. The underlying theory is that cognitive processes, especially those requiring significant mental effort or attention, can increase pupil size. This phenomenon is supported by the PLRT, which posits that pupil dilation is not solely a response to changes in light, but it can also be influenced by CL and emotional states. This theory is grounded in neuroscientific research that links activity in the Locus Coeruleus-Norepinephrine (LC-NE) system with pupil dilation attention and arousal processes (Unsworth & Robison, 2016).

2.2 Empirical Background

This section provides an exhaustive overview of the existing scholarly literature on four key domains related to CL research, namely methodological approaches to measuring CL, the role of learning modalities and individual differences in CL, and the relationship between PATs and CL. The latter will be followed with exploring their beneficial aspects as well as drawbacks.

There are various subjective measures of CL (e.g., Klepsch et al., 2017) were developed in connection with the first domain; however, the validity and reliability of self-reported measures such as the NASA Task Load Index and Paas Cognitive Load Scale have been questioned due to subjectivity, complexity of CL, context dependency, and interpretation of scores. Besides, Chen and Epps (2014) noted that the effectiveness of pupillometry in measuring CL may be influenced by the type and level of the task, with pupil diameter, blink rates, and gaze patterns as the most appropriate indicators of CL. Hebbar et al. (2022) reported that Task-Evoked Pupillary Responses (TEPRs) can measure CL in VR training when corrected for light reflexes. They add that light reflexes allow for real-time assessment and instructional design improvements in VR training scenarios. TEPRs have been validated to predict cognitive load accurately in VR settings, offering insights into the cognitive demands of tasks performed in virtual environments.

Regarding the second domain, Lin et al. (2016) found that Chinese EFL learners showed a positive association between the number of engaged learning channels and successful learning, unlike native English speakers who experienced split attention and high CL. Similarly, Wagner-Loera (2018) integrated a Reduced Cognitive Load Classroom (RCLC) approach with flipped teaching, creating a low-distraction environment that led to effective learning and satisfied learners. Furthermore, Song et al. (2023) found that optimizing a VR environment for Korean EFL learners led to enhanced learning due to reduced cognitive load. Moreover, learning style is a critical factor, with auditory learners potentially experiencing a higher cognitive load in multimedia learning than visual or kinesthetic learners (Çakiroğlu et al., 2020).

Considering the last domain, research has demonstrated potential complex relationships between the levels of CL and PATs, concentrating on practicing voiced vs. voiceless sounds. Lewis et al. (2010) questioned the direct relation between phonological awareness skills and speech perception in noise, indicating that other factors, like environmental factors, may play a role. Besides, Chiu et al. (2019) demonstrated that CL can negatively impact speech perception, increasing reliance on lexical knowledge and decreasing reliance on phonetic detail. Moreover, the type of phoneme used in PA tasks can significantly impact performance, with specific phonemes leading to better outcomes (Gabrić & Vandek, 2021). Finally, Higher working memory capacity was correlated with more native-like phonological processing in L2 (Darcy et al., 2015).

The positive and negative aspects observed in extant literature about similar studies are herein delineated to offer a comprehensive overview and illuminate the significance and potential of the present study. While the reviewed literature has yielded valuable insights into the intricacies of various language acquisition processes and assessed their efficacy or limitations in terms of educational outcomes, the corpus of research primarily focusing on the cognitive dimensions of reading and phonological processing concerning the levels of CL experienced by English learners remains notably sparse. Notably, studies exploring CL have predominantly relied on the contextual validity and reliability of existing assessment tools, thereby engendering debates due to the oversight of confounding variables such as learners' demographic characteristics, language proficiency, cognitive aptitudes, and overall health status. Conversely, There is a paucity of research regarding the use of more robust measurement instruments and methodologies, such as VR cognitive pupillometry analysis, for assessing CL. Moreover, the existing literature exclusively originates from non-Iranian English learning environments, necessitating further investigations within the Iranian educational milieu to elucidate the potential impacts of instructional modalities or the incorporation of challenging phonetic elements from Persian on learners' CL levels. Lastly, while prior studies have predominantly focused on the effects of language-learning interventions on cognitive processes during instruction, limited attention has been directed toward identifying underlying causes and proposing actionable solutions for educators, instructional material developers, and curriculum designers. In light of the gaps mentioned above in the literature, the current study attempted to provide answers to the following research questions.

1. Is there any significant difference between the CL levels encountered by the Iranian pre-intermediate EFL learners while performing PATs emphasizing voiced vs. voiceless 'th' sounds?
2. Is there any significant difference between the CL levels experienced by the Iranian pre-intermediate EFL learners performing in auditory vs. auditory/visual PATs focusing on 'th' sound variations?
3. What are the potential causes of the varying levels of CL measured in PATs, considering the sound variations and delivery modes?

3. Method

3.1 Design

A sequential explanatory research design was selected to increase the validity and reliability of the study outcomes. This approach involved two distinct phases: first, quantitative data is collected and analyzed, followed by the collection and analysis of qualitative data based on the quantitative results. This design aims to use qualitative data to elucidate the quantitative findings. In the quantitative phase, the participants were divided into four equal groups: G1 and G2 (to address RQ1) and G3 and G4 (to address RQ2). The participants in G1 experienced auditory/visual PATs practicing voiced "th" (ð) or 'eth', while those in G2 received auditory/visual PATs focusing on voiceless "th" (θ) or "theta". The



remaining participants were divided into two additional groups, G3 and G4, where the former experienced auditory PATs and the latter faced auditory/visual PATs, concentrating on both variations of "th" sounds. In the qualitative phase (to address RQ3), an equal number of participants were selected from all groups to explore the potential reasons behind the varied levels of CL experienced due to the sound variations (G1 and G2) and modes of PAT delivery (G3 and G4). This study lasted from December 2023 to March 2024 and was conducted in a private Speech Language Pathology (SLP) Center in Tehran. There were two main reasons for selecting this center: first, its highly advanced VR equipment, including VR simulators and fully immersive VR-HMDs, and second, its local network platform for simultaneously VR-authoring the required instructional material and measuring the CL of the learners.

3.2 Participants

The study employed a rigorous participant selection process to ensure the homogeneity of the sample and minimize potential confounding factors that could influence the levels of encountered CL among the participants. The participants were selected to be of the same gender to control for the potential biological influences on CL. Furthermore, the age range of the participants was narrowly constrained, with a minimum age of 15 and a maximum age of 17, to minimize the impact of developmental differences on the study outcomes. According to schema theory and CLT, participants with considerable previous exposure to PATs are likely to experience less CL during the study than their peers with less or no exposure to PATs. Therefore, the study sought to select participants with similar English proficiency levels, specifically pre-intermediate EFL learners. Due to time constraints, a general English placement test could not be conducted. Instead, the participants were selected from those who had successfully passed the subsequent oral and written tests following their pre-intermediate English courses. Additionally, the researchers removed any participants whose final scores significantly deviated from the mean, as determined by calculating the interquartile range (IQR) to ensure homogeneity further. After the initial participant selection and removing outliers, the final sample consisted of 36 male pre-intermediate EFL learners divided into four groups: G1, G2, G3, and G4. Table 1 provides a detailed overview of the characteristics of the participants.

Table 1- Characteristics of the participants

Groups	Focus	Frequency	Gender	Intervention
G1	Sound variations	9	M	Auditory and visual PATs on ð
G2		9	M	Auditory and visual PATs on θ
G3	PATs delivery mode	9	M	Only-auditory PATs on ð & θ
G4		9	M	Auditory and visual PATs on ð & θ

According to Table 1, the participants in the G1 and G2 experienced the same auditory and visual PATs. However, in the former, the focus was on the voiced version of the 'th' sound, and in the latter, the concentration was on the voiceless subsequent sound. The participants in the G3 and G4 received the same PATs, focusing on both versions of "th" sounds but across two different modes of delivery in which the former received auditory types of PATs and the latter underwent auditory/visual PATs. During the conduction of this study, the participants were ensured that their personal information was kept confidential.

3.3 Instruments

3.3.1 Tobii Nexus

Tobii Nexus is a cutting-edge software that integrates eye-tracking technology into devices with webcams. It enhances user experiences by detecting presence, attention, and intent, providing real-time data streams like gaze point and pupil diameter. Advanced pupil detection algorithms (i.e., Starburst, Swirski, ExCuSe, ElSe, PuRe, and PuReST) can be used by utilizing other software to ensure precise measurements of cognitive pupillometry. The primary goal of Tobii Nexus is to provide a comprehensive solution for conducting accurate and reproducible pupil examinations, positioning itself as a competitor to high-end commercial systems in the market. This study used the Swirski algorithm (Swirski et al., 2012) due to its robustness and accuracy compared to other analyzing algorithms. It includes 11 parameters, which is more than the other algorithms, allowing for greater flexibility in adjusting the processing parameters to optimize the pupil contour's detection accuracy. Additionally, the Swirski algorithm is licensed under MIT, providing more freedom in its experimental use than the other algorithms licensed for particular uses only.

3.3.2 VR-HMD

A VR-HMD, which stands for Virtual Reality Head-Mounted Display, is a wearable device that presents visuals directly to the eyes, creating an immersive virtual reality experience. VR headsets are HMDs combined with IMUs (Inertial Measurement Units) to enhance the virtual reality experience. The technology behind VR-HMDs involves optics to fill the user's entire field of vision, head tracking technology for accurate positioning data, and other hardware components for comfort and functionality. This study used HTC Vive Pro 2, a high-end VR-HMD known for its high visual fidelity, balanced ergonomics, and sub-millimeter tracking accuracy. This study used specific strategies to control lighting conditions, ambient noise levels, and distractions in the surroundings to avoid their potential confounding roles in measuring CL. Moreover, a series of guidelines were followed to ensure that participants did not experience severe



levels of VR or motion sickness as much as possible, including (a) using high-quality VR equipment with low latency and high refresh rates, (b) avoiding constant moving of head or eyes, (c) doing real-time monetization of the user by an observer, (d) designing microlearning nuggets of the material which are concise, focused learning units that break down complex subjects into easily digestible portions, enhancing learning retention and engagement.

3.4 Data Collection Procedure

3.4.1 Pre-intervention Phase

In the pre-intervention phase, specific steps were followed to collect the required data. First, the required permissions concerning the conduction of this study were received from the head of a private SLP center and the manager of a language institute in Tehran. Second, after two different attempts, 36 male EFL learners with pre-intermediate English proficiency were selected. Due to time constraints, instead of checking the participants' English proficiency using a placement test, participants were selected based on their pass/fail final score index in the pre-intermediate English course tests. Besides, the ones that were determined as outliers were omitted. Moreover, the age range was two years old, which was acceptable as age is considered a potential factor that may fluctuate the experienced levels of CL among the participants. As the age of the final participants was lower than 18 years old, it was necessary to inform one of the parents of their child's participation in this study. Third, the auditory PATs, including (a) rhyme recognition and (b) sound isolation and auditory/visual PATs consisted of (a) picture sorting and (b) word discrimination focusing on the voiced and voiceless version of "th" sounds were designed to be appropriate for VR-authoring. Fourth, the participants were randomly divided into four groups; for the G1 and G2 groups, the focus was on measuring CL regarding performing PATs concentrating on voiced vs. voiceless "th" sounds, and for the G3 and G4, CL was measured with a focus on PATs modes of delivery that was auditory vs. auditory/visual channels.

3.4.2 Intervention Phase

During the intervention phase, a sequence of actions was carried out in the fifth step. The procedures for carrying out each of them are described here to understand better each PAT and how CL was measured. In the (a) rhyme recognition as the first auditory PAT, participants with VR-HMDs were provided an audio clip of a word containing the voiced or unvoiced "th" sound. Then, they were asked to repeat the word and identify the "th" sound. After the participants had identified the "th" sound, an audio clip that included a list of rhyming words, including the target word, was played. Then, they were required to identify the word that rhymed with the target word. These processes were repeated four times but with different words. At the same time, participants watched a blank white screen to avoid imposing extra cognitive load, as white is reported to be a neutral color concerning its effect on memory. The participants' pupil dilations, blink rates, and gaze patterns were measured by the Swirski algorithm with the Tobii Nexus software while performing the mentioned PAT four times. The output of these indexes was reported as total mean scores.

In the (b) sound isolation as the second auditory PAT, participants in a similar condition to the rhyme recognition mentioned above were provided with an audio clip that included the words that had voiced or unvoiced "th" sounds in different positions including beginning, middle, and at the end parts. Then, they were prompted to isolate the "th" sound from different parts of the words, mention their places, and repeat it. The repeated frequency of the process and the way CL was measured were similar to the previous PAT. In the (a) picture sorting, as the first auditory/visual PAT, the participants were shown different pictures on one screen, and their subsequent written presentations were below each picture. They used a joystick to grab each word with a voiced or unvoiced "th" sound and put it in a circle shape or a squared box in the appropriate place. The number of PAT repetitions and the procedure for measuring CL were the same as the previous PATs. In the (b) word discrimination, as the second auditory/visual PAT, a series of words in written forms were presented on the screen across different pairs, and the participants were needed to determine which word in each pair included voiced or unvoiced "th" sounds with the use of their joysticks. The PAT and the CL evaluation process frequency was the same as other PATs. Table 2 illustrates the characteristics of each PAT.

Table 2- Characteristics of PATs

PATs	Type of PATs	Procedure
Auditory	Rhyme recognition	For instance, the participants listened to the word "bath" and identified the "th" sound. Then they heard the list of rhyming words (e.g., path, math, wrath, lath) and selected "path" as the word that rhymed with "bath".
	Sound isolation	For example, the participants heard each word and isolated the "th" sound, stating its position in the word (beginning like think, middle like a brother, or end like a tooth). For instance, for "think" they said, "the 'th' sound was at the beginning of the word".
Auditory/ visual	Picture sorting	The participant saw pictures like a bath, math, path, and wreath. Below each picture was the written word. Using a joystick, they grabbed the word with the "th" sound (like "path" or "wreath") and placed it in the appropriate shape (circle or square).
	Word discrimination	Using the joystick, they selected the word in each pair (bath/bat-moth/mop-breath/bread) that contained the "th" sound, such as "bath", "moth", and "breath".

3.4.3 Post-intervention Phase

In step six, after the intervention phase, to have a comprehensive insight into the outcomes of the quantitative phase and the possible reasons behind them, 8 participants were randomly selected from all of the groups to participate in a series of peer-to-peer structured interviews in which they were asked "how mentally demanding was the task?" and "what do you think were the reason(s)?" These items were borrowed from the "mental demand" section of NASA TLX, an outdated but context-specific reliable questionnaire to examine cognitive load in scientific studies. It should be mentioned that the participants had pre-intermediate English proficiency; they were allowed to switch to Persian in cases where they could not fully transfer necessary information.

3.5 Data Analysis Procedure

Before answering the research questions, it was necessary to check a series of statistical assumptions to determine whether to use parametric or non-parametric statistical tests. The assumptions included (a) ensuring independence of observations (no relationship or correlation between the observations in one group and those in the other group), (b) checking if the collected data was approximately normally distributed (Shapiro-Wilk test), (c) verifying the homogeneity of the variances (Levene's test). As all assumptions were met, descriptive statistics and three independent sample t-tests were computed for each of the first and second research questions. The measuring of CL was done through cognitive pupillometry analysis and compared by calculating three total mean scores for (a) frequencies of pupil diameter compared to the average pupil size in normal conditions, (b) frequencies of blinks or blink rates, and (c) gaze patterns in terms of fixation duration frequencies. A gaze pattern is the total time spent fixating on a point before moving the eyes to another point, known as a saccade. A thematic analysis was carried out to answer the third research question regarding the influential factors in determining the levels of mental demand for participants in each group.

4. Results

4.1 Quantitative Phase: RQ1 Outcomes

It was ensured that the required statistical assumptions were met before answering the first research question. Regarding pupil diameter, descriptive statistics showed a mean value of 5.84 for the participants in the G1 (δ), while for the ones in the G2 (θ), the mean value was 5.77. The results showed that participants in both groups experienced dilated pupils as the mean values were between 4 and 8 millimeters (Koch et al. 1991). Thus, it can be concluded that both of these groups faced increased pupil diameter, inferring the experience of great mental effort and cognitive processing leading to the experience of high CL ($PD \geq 4mm$) (Gavas et al., 2017; Kiefer et al., 2016). Concerning the blink frequency per minute, the mean value for the participants in the G1 (δ) was 11.3, while for the ones in the G2 (θ), it was 12.6. It can be concluded that the blink frequency was lower than in the normal situation ($BR \geq 17$), justifying the presence of high perceptual load, inferring the experience of high CL levels among the participants (Bentivoglio et al., 1997; Chen & Epps, 2014; Gowrisankaran et al., 2012). With a focus on gaze pattern, fixation durations were calculated in terms of two total mean values for the participants in G1 (δ), 527 milliseconds, and G2 (θ), 494 milliseconds. The average fixation duration typically falls between 150 and 300 (Galley et al., 2015), and the calculated mean values show increased CL, described as experiencing moderate CL levels (Negi & Mitra, 2020). To better understand, Figure 1, based on 100% stacked columns, shows the insignificant differences between G1 and G2 regarding pupil diameter, blink frequency, and gaze pattern.

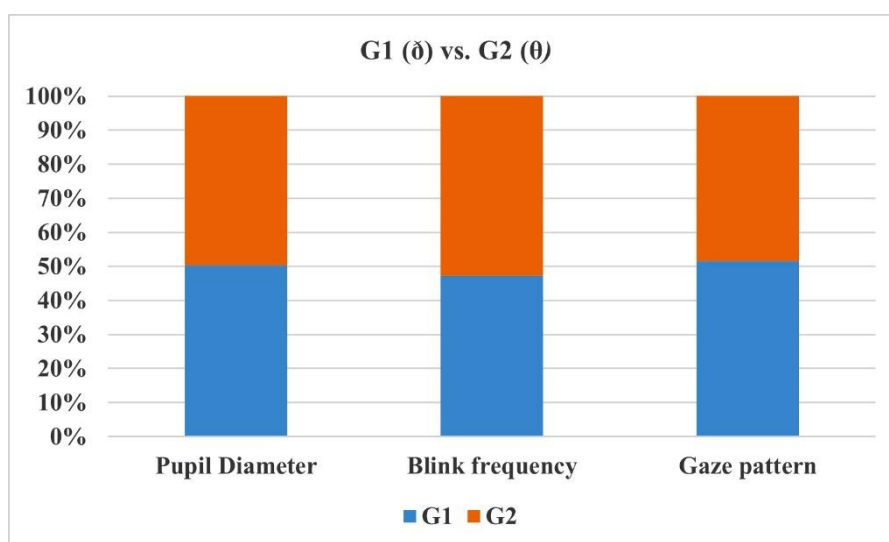


Figure 1- Percentage Stacked Columns Contrasting G1 (δ) vs G2 (θ)

Table 3 indicates the results of three independent sample t-tests across three measures of CL among G1 and G2.

Table 3- Multiple Independent Sample T-tests for CL Measures Between G1 and G2

	t	Sig.(2tailed)	M Difference	Std. Error Diff.	Lower	Upper
Pupil diameter	.466	.648	.186	.400	-.662	2.771
Blink rates	.648	.504	.333	.487	-.700	1.336
Gaze pattern	.394	.699	.222	.563	-.973	1.417

The participants in both groups experienced increased levels of CL based on the results of pupil diameter (high levels of CL), blink rates (high levels of CL), and gaze pattern or fixation duration (moderate levels of CL). As seen in Table 2 (all the p-values are more than the significance 0.05 level), there was no significant difference between the CL levels experienced by the Iranian pre-intermediate EFL learners while performing PATs focusing on voiced vs. voiceless 'th' sounds regarding the mentioned three measures.

4.2 Quantitative Phase: RQ2 Outcomes

Before addressing the second research question, the researchers undertook a comprehensive evaluation to ensure that the requisite statistical assumptions were thoroughly satisfied. Descriptive statistics revealed that participants in Group 3 (auditory mode) had an average pupil diameter of 6.12, while those in Group 4 (auditory/visual mode) had an average of 5.68 millimeters. The outcomes demonstrated that participants in both groups experienced dilated pupils as the mean values were posited between 4 to 8 millimeters (Koch et al. 1991), suggesting the experience of heightened mental effort and high CL levels ($PD \geq 4mm$) (Gavas et al., 2017; Kiefer et al., 2016). Additionally, the blink frequency per minute was lower than average ($BR \geq 17$) for both groups ($G3 = 10.8$ and $G4 = 13.1$), indicating a high perceptual load and increased levels of CL among participants (Bentivoglio et al., 1997; Chen & Epps, 2014; Gowrisankaran et al., 2012). With a focus on gaze patterns, fixation durations were computed regarding two total mean values for the participants in G3 (verbal), 598 milliseconds, and G4 (verbal/visual), 514 milliseconds. As the standard range for fixation duration is between 150 to 300 milliseconds (Galley et al., 2015), the computed mean values revealed an increased amount of CL, which is classified as having moderate levels of CL (Negi & Mitra, 2020). As illustrated in Figure 2, a 100% stacked column chart is employed to elucidate the significant disparities in pupil diameter, blink frequency, and gaze patterns between G3 and G4.

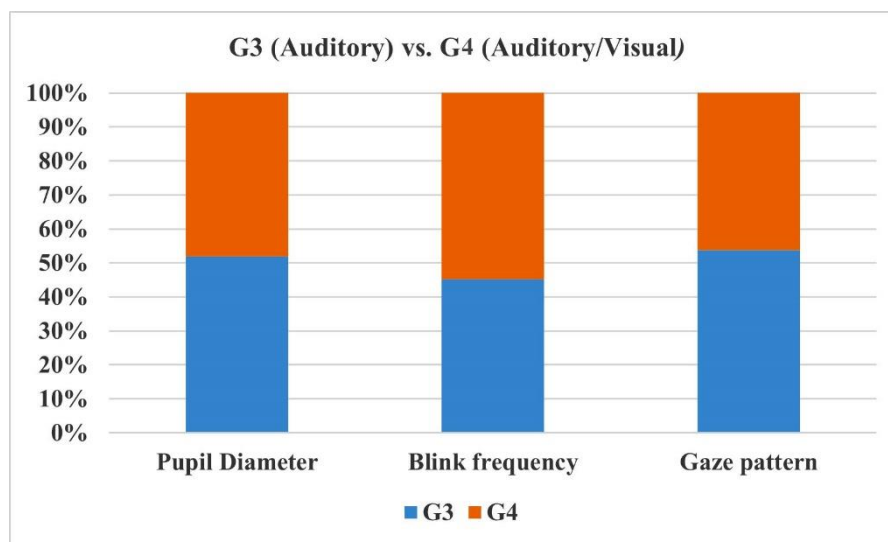


Figure 2- Percentage Stacked Columns Contrasting G3 (Auditory) vs. G4 (Auditory/Visual)

The outcomes of three independent sample t-tests across three domains of CL among G3 and G4 are shown in Table 4.

Table 4- Multiple Independent Sample T-tests for CL Measures Between G3 and G4

	T	Sig.(2tailed)	M Difference	Std. Error Diff.	Lower	Upper
Pupil diameter	.867	.003	-.566	.641	-1.914	.804
Blink rates	.544	.001	.791	-.111	.412	.762
Gaze pattern	.270	<.001	.487	.126	-.985	.559

The participants in both groups experienced elevated levels of CL based on the results of pupil diameter classified as high levels of CL, blink rates categorized as high CL levels, and gaze pattern or fixation duration determined as moderate CL levels. As shown in Table 3 (all the p-values are less than the 0.05 critical level), there was a significant difference between the CL levels experienced by the Iranian pre-intermediate EFL learners performing in auditory vs. auditory/visual PATs focusing on 'th' sound variations. It can be concluded that the participants in the auditory PATs group (G3) had experienced significantly higher levels of CL compared to their peers in the verbal/visual PATs group.

4.3 Qualitative Phase: RQ3 Outcomes

Concerning answering RQ3, which was associated with the qualitative phase of this study, A series of peer-to-peer interviews were carried out with eight randomly selected participants from all groups (two selected from each) to determine the potential reasons behind the revealed levels of experiencing CL in PATs, considering the sound variations (G1 and G2) and delivery modes (G3 and G4). Two items were asked of them: first, how they evaluated the tasks regarding their mental demands and their opinions on the potential reasons behind the levels of CL which they experienced. The findings of the RQ3 demonstrated that all of the interviewees pointed out that the PATs required high mental demands and attention, confirming the results of the quantitative phase. Considering the potential reasons, the absence of the 'th' sound in Persian or lack of frequent exposure in everyday communication was mentioned as one reason for the high complexity of the PATs perceived by the participants. For instance, interviewee number 3 mentioned that

"In my opinion, the absence of the 'th' sounds in Persian was the leading factor that made PATs considerably tricky and challenged me in distinguishing its voiced and voiceless variations. If I had frequent exposure or previous familiarity with these sounds, I would have had a significantly less complicated time determining and manipulating them."

Another factor the interviewees stated to have a considerable influence on the difficulty of performing PATs was the motor control required to move articulatory organs regarding 'th' sound variations. For example, interviewee number 5 pinpointed that

"I suppose that if the PATs were structured around other sounds in English, considerably less attention would be needed, and performing them would be automatic. Positioning my tongue lightly against the upper front teeth was complicated. It needed instantaneous coordination between mind and tongue. So, most of the time, I was predisposed to substitute similar 'th' sounds with 'S' or 'Z'!"

Incorrect frequent use and manipulation of the 'th' sounds by the interviewees made it considerably challenging for them to rewire their internalized system of English pronunciation. As an example, interviewee number 7 stated that

"On the one hand, since I started learning English, I tended to pronounce voiced and voiceless 'th' sounds not in an interdental correct form, which made it significantly complicated for me to deal with the PATs. On the other hand, reshaping the fixed behavior of wrongly manipulating 'th' sound variation was not that easy."

Difficulties in memorizing phonological patterns while performing PATs, notably the verbal ones, were another main influential factor mentioned by the interviewees, and they faced high mental demands. For instance, interviewee number 4 reported that

"I experienced higher difficulty levels for auditory PATs than auditory/visual PATs, as they required me to memorize and recall a series of particular phonological patterns. Simultaneously recalling challenging variations of 'th' sounds while harmonizing my mind and tongue to produce the appropriate one for different types of PATs was considerably troublesome."

One of the other variables that influenced the interviewees' performance while performing PATs was their inability to shift between various auditory, visual, and mental tasks quickly or simultaneously. For instance, interviewee number 6 highlighted that

"It was a complex undertaking to simultaneously listen to the challenging 'th' sounds while determining them in various words and matching them in auditory/visual PATs. In most cases, for auditory/visual PATs, paying attention to both visual and auditory channels was demanding. It was like managing two engines in one car simultaneously."

The last influential variable that affected the perceived imposed level of mental demand concerning performing PATs was their delivery types as in the auditory/visual PATs; auditory information was supported by visual cues, leading to better performance. For example, interviewee number 1 pointed out that

"I believe the more sensory input engaged in PATs, the better the learners' performance would be. In cases where auditory and visual information were presented, such as auditory/visual PATs, I had considerably fewer problems correctly performing the tasks. In contrast, more time was needed to handle the task in just verbal ones."

The thematic analysis showed six distinct influential factors that had significant roles in imposing various levels of CL on the participants while performing PATs.

5. Discussion

The results of RQ1 showed that while participants encountered high levels of CL, including dilated pupils, high blink frequency, and increased gaze pattern or fixation duration, no significant difference was found between them concerning the encountered CL levels across 'th' sound variations. On the one hand, the result aligns with that of Jensen and Thøgersen (2017), in which the features of foreign accent speech and foreign language phonology engage the brain of non-native language learners with considerably more cognitive processes in all situations, leading to increased levels of experiencing CL. On the other hand, the RQ1 result does not agree with that of Soleimani and Rezazadeh (2014),

who found that an increase in task cognitive complexity led to greater accuracy and linguistic complexity among Iranian EFL learners. The potential reason behind the inconsistency of the RQ1 result with the mentioned study may be the type and characteristics of linguistic activity.

The outcomes of RQ2 shed light on the fact that while participants faced high levels of CL, including increased pupil diameter, elevated blink rate, and increased fixation duration or gaze pattern, a significant difference was found between them concerning the experienced CL levels across PATs modes of delivery as the ones in auditory PATs encountered significantly higher CL compared to their peers in auditory/visual PATs. The results of RQ2 are in harmony with that of [Lin et al. \(2016\)](#), who reported that engaging multiple sensory channels, as mentioned in the DCT while learning a foreign language, decreased the amount of CL experienced by the learners. Also, the result of the RQ2 contradicts that of [Brünken et al. \(2004\)](#), who argued that the audiovisual presentation of auditory and pictorial learning materials increased the demand on phonological cognitive capacities, confirming the modality effect.

The debate between DCT ([Mayer & Pilegard, 2005](#)) and the split attention effect has significant implications for instructional design and learning strategies. DCT ([Kalyuga et al., 1999](#)) argues that presenting information in both verbal and visual formats can enhance learning outcomes. On the other hand, the split attention effect cautions against presenting information in a way that forces learners to split their attention, which can hinder comprehension and retention. These differing perspectives highlight the need for further research to understand the most effective instructional design strategies. Finally, RQ3 was proposed to explore the potential reasons behind the varied levels of CL experienced in PATs, considering the sound variations and modes of delivery. The findings revealed that all PATs were mentally demanding for the Iranian pre-intermediate EFL learners. The RQ3 finding follows that of [Shalabi \(2017\)](#), who reported that Arab, Chinese, and Pakistani EFL learners faced considerable mental demands and cognitive challenges while focusing on voiced vs. voiceless 'th' sounds.

Moreover, the findings or RQ3 showed six variables that were reported to influence the levels of CL experienced by the Iranian pre-intermediate EFL learners while performing PATs, including (a) segmental differences (referring to the absence of voiced and voiceless 'th' sounds in Persian), (b) phonological transfer (referring to the transfer of 's' and 'z' sound from Persian into English instead of the correct pronunciation of interdental fricatives such as voiced and unvoiced 'th' sounds), (c) phonological fossilization (referring to as the inability to change the incorrect internalized phonological behavior of pronouncing 'th' sound variations), (d) working memory constraints (referring to the varied levels of working memory capacity among individuals while performing PATs), (e) cognitive flexibility (referring to the various ability of individuals in quickly or simultaneously shift between different mental tasks including the visual and auditory ones), and (f) task delivery (referring to the design of the instructional materials and tasks and the number and amounts of sensory inputs involved in them). There are various reasons behind the findings, which can be justified from different points of view, such as linguistic, cognitive, and material development described in the following.

Research suggests that the absence of certain sounds in a learner's first language (L1) can impact their ability to learn those sounds in a second language (L2), particularly in phonology and sound patterns. However, explicit phonetics instruction has improved L2 learners' perception of these sounds ([Kissling, 2015](#)). In addition, [Dijkstra et al. \(1999\)](#) found that phonological contrasts between languages can have both facilitatory and inhibitory effects on word recognition, suggesting that the impact of interlanguage phonology on cognitive load may depend on the specific linguistic context, which needs more studies to be conducted. Moreover, children with specific language impairments may be more prone to phonological fossilization and have diminished phonological working memory capacity, which can compromise their correct pronunciation and sentence comprehension efforts. Also, working memory capacity and the phonological loop in pronunciation tasks play crucial roles, particularly in language acquisition and processing. [Baddeley et al. \(1998\)](#) emphasized the importance of the phonological loop and muscle memory in learning new phonological forms and acquiring novel phonological and grammatical structures, which was supported by [Wilson and Emmorey \(1997\)](#), who found evidence for a visuospatial "phonological loop" in working memory, suggesting that working memory can develop a language-based rehearsal loop in the visuospatial modality. Finally, cognitive flexibility is involved in assembling phonological representations from orthographic inputs, with the left inferior prefrontal cortex and bilateral parietal cortices being key areas ([Clark & Wagner, 2003](#)). This ability differs among language learners and considerably influences their perception of the mental loads of phonological tasks ([Llompart & Reinisch, 2019](#)).

6. Conclusion

This study developed valuable insights into cognitive linguistics, particularly cognitive phonology, by revealing that Iranian pre-intermediate EFL learners encountered high levels of CL regarding two measures of pupil diameter and blink frequency and moderate levels concerning the gaze pattern or fixation durations while performing PATs regarding voiced and unvoiced 'th' sounds across auditory and auditory/visual modes of task delivery which in general shed light on the high levels of total CL they experienced. Also, while the sound variations did not lead to significant differences among the Iranian pre-intermediate EFL learners concerning the perceived high levels of CL, the types of PATs made considerable differences in which auditory PATs imposed higher CL on the EFL learners' cognition compared to verbal/visual ones. Furthermore, it was found that segmental differences, phonological transfer, phonological fossilization, working memory, cognitive flexibility, and task delivery were the possible causes that affected the experienced levels of CL among Iranian pre-intermediate EFL learners, but further investigations are still needed. This study provided practical guidelines for educators and researchers in designing effective teaching strategies that optimize



learning outcomes. Also, this study demonstrated the potential physiological measures of VR cognitive pupillometry analysis as reliable indicators of CL during language learning tasks, especially phonological ones, leading to methodological advancements in cognitive linguistics. This study proposes future studies that follow the same objectives but with a focus on English speakers with diverse proficiency. In addition, utilizing different advanced measures of cognitive load such as heart rate, skin conductance, breathing rate, and heart rate variability provides valuable outcomes and critically reveals the hidden cognitive aspects of language learning and performing language tasks.

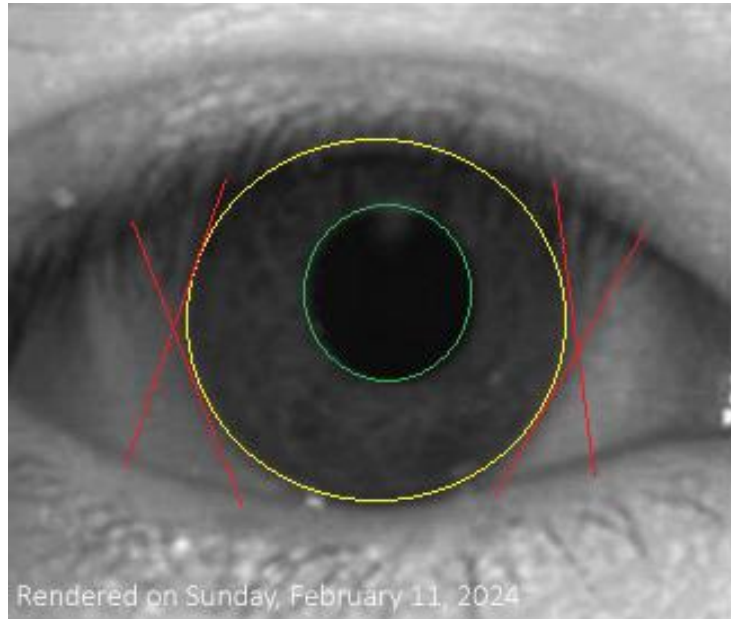
References

- Alswede, K. (2022). *Measuring the effect of phonemic awareness intervention on students' fluency levels*. Elmhurst University Publication.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review* 105(1), 158-169.
- Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders* 12(6), 1028-1034.
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science* 32(1), 115-132.
- Burns, N. (2023). *Inspiring deep learning with metacognition: A guide for secondary teaching*. Sage.
- Çakiroğlu, Ü., Güler, M., Atabay, M., & Güler, M. (2020). Connections between learning styles and perceived cognitive load in multimedia learning: An experimental study. *Journal of Educational Technology Systems* 48(4), 553-573.
- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read?. *Cognition* 91(1), 77-111.
- Chen, S., & Epps, J. (2014). Using task-induced pupil diameter and blink rate to infer cognitive load. *Human-Computer Interaction* 29(4), 390-413.
- Chiu, F., Rakusen, L. L., & Mattys, S. L. (2019). Cognitive load elevates discrimination thresholds of duration, intensity, and f for a synthesized vowel. *The Journal of the Acoustical Society of America* 146(2), 1077-1084.
- Clark, D., & Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia* 41(3), 304-317.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 108(1), 204-216.
- Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences* 40(1), 63-72.
- Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language* 41(4), 496-518.
- Djiguimkoudre, N. (2021). Assessing the relevance of phonemic awareness activities in foreign language teaching and learning. *American Scientific Research Journal for Engineering, Technology, and Sciences* 75(1), 45-69.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly* 36(3), 250-287.
- Fabre-Merchan, P., Torres-Jara, G., Andrade-Dominguez, F., Ortiz-Zurita, M. J., & Alvarez-Muñoz, P. (2017). A phenomenological study: The impacts of developing phonetic awareness through technological resources on English language learners' communicative competences. *English Language Teaching* 10(12), 83-87.
- Gabrić, P., & Vandek, M. (2021). Striking differences in performance patterns and neuropsychological correlates within semantic and phonemic fluency tasks. *PsyArXiv* 15(5), 65-78.
- Galley, N., Betz, D., & Biniossek, C. (2015). Fixation durations: Why are they so highly variable?. *Das Ende von Rational Choice? Zur Leistungsfähigkeit der Rational Choice Theorie* 93(1), 1-26.
- Gavas, R., Chatterjee, D., & Sinha, A. (Eds.). (2017). *Estimation of cognitive load based on the pupil size dilation. Systems, man, and cybernetics*. IEEE Publication.
- Gowrisankaran, S., Nahar, N. K., Hayes, J. R., & Sheedy, J. E. (2012). Asthenopia and blink rate under visual and cognitive loads. *Optometry and Vision Science* 89(1), 97-104.
- Griffith, P. L., & Olson, M. W. (1992). Phonemic awareness helps beginning readers break the code. *The Reading Teacher* 45(7), 516-523.
- Hebbar, A., Vinod, S., Shah, A. K., Pashilkar, A., & Biswas, P. (2022). Cognitive load estimation in VR flight simulator. *Journal of Eye Movement Research* 15(3), 25-39.
- Jensen, C., & Thøgersen, J. (2017). Foreign accent, cognitive load and intelligibility of EMI lectures. *Nordic Journal of English Studies*, 16(3), 107-137.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 13(4), 351-371.
- Kanokpermpoon, M. (2013). Managing working memory in language instructions: An overview of cognitive load



- theory. *Thammasat Review* 16(2), 93-108.
- Kiefer, P., Giannopoulos, I., Duchowski, A., & Raubal, M. (Eds.). (2016). *Measuring cognitive load for map tasks through pupil diameter. Geographic information science*. Springer.
- Kissling, E. M. (2015). Phonetics instruction improves learners' perception of L2 sounds. *Language Teaching Research* 19(3), 254-275.
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology* 8(1), 19-35.
- Koch, D. D., Samuelson, S. W., Haft, E. A., & Merin, L. M. (1991). Pupillary and responsiveness: Implications for selection of a bifocal intraocular lens. *Ophthalmology*, 98(7), 1030-1035.
- Lewis, D., Hoover, B., Choi, S., & Stelmachowicz, P. (2010). The relationship between speech perception in noise and phonological awareness skills for children with normal hearing. *Ear and Hearing* 31(6), 761-772.
- Lin, J. J., Lee, Y. H., Wang, D. Y., & Lin, S. S. (2016). Reading subtitles and taking notes while learning scientific materials in a multimedia environment: Cognitive load perspectives on EFL students. *Journal of Educational Technology & Society* 19(4), 47-58.
- Llompert, M., & Reinisch, E. (2019). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition* 22(5), 1085-1100.
- Mayer, R. E., & Pilegard, C. (2005). *Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles*. Cambridge.
- Negi, S., & Mitra, R. (2020). Fixation duration and the learning process: An eye tracking study with subtitled videos. *Journal of Eye Movement Research* 13(6), 56-69.
- Rokhman, M. F., Lintangari, A. P., & Perdhani, W. C. (2020). EFL learners' phonemic awareness: A correlation between English phoneme identification skill toward word processing. *Journal of English Educators Society* 5(2), 135-141.
- Shalabi, M. (2017). Think, tink or sink, the phonological awareness of English voiceless interdental fricative [θ] and [ð] among Chinese, Arab and Pakistani learners of English. *American Journal of Education and Information Technology* 1(3), 31-37.
- Smalle, E. H., Muylle, M., Duyck, W., & Szmalec, A. (2021). Less is more: Depleting cognitive resources enhances language learning abilities in adults. *Journal of Experimental Psychology: General* 150(12), 24-33.
- Soleimani, H., & Rezazadeh, M. (2014). The effect of increase in task cognitive complexity on Iranian EFL learners' accuracy and linguistic complexity: A test of Robinson's Cognition Hypothesis. *Applied Research on English Language* 3(1), 41-54.
- Song, C., Shin, S. Y., & Shin, K. S. (2023). Optimizing foreign language learning in virtual reality: A comprehensive theoretical framework based on constructivism and cognitive load theory. *Applied Sciences* 13(23), 125-136.
- Swirski, L., Bulling, A., & Dodgson, N. (2012). Robust real-time pupil tracking in highly off-axis images. *Eye Tracking Research and Applications* 24(3), 173-176.
- Tan, K. H., Wheldall, K., Madelaine, A., & Lee, L. W. (2007). A review of the simple view of reading: Decoding and linguistic comprehension skills of low- progress readers. *Australian Journal of Learning Difficulties* 12(1), 19-30.
- Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience* 16(1), 601-615.
- Wagner-Loera, D. (2018). Flipping the ESL/EFL classroom to reduce cognitive load: A new way of organizing your classroom. In J. Mehring, & A. Leis (Eds.), *Innovations in flipping the language classroom* (pp. 169-184). Springer.
- Wilson, M., & Emmorey, K. (1997). A visuospatial "phonological loop" in working memory: Evidence from American Sign Language. *Memory & Cognition* 25(3), 313-320.

Appendix A. Normal pupil



Appendix B. Dilated Pupil

