University of Tehran press

JOURNAL OF FOREIGN LANGUAGE RESEARCH
p-ISSN:2588-4123 e-ISSN:2588-7521
https://jflr.ut.ac.ir **Email**:jflr@ut.ac.ir

Investigating the Impact of Visual Feedback on Voice Onset Time (VOT) among Iranian EFL Learners

Forough Amirjalili* 0000-0002-6149-0416
Department of English Language and Literature, Yazd University, Yazd, Iran
Email: f.amirjalili@yazd.ac.ir

ABSTRACT
This study delved into the effectiveness of Visual Feedback (VF) as a tool for enhancing the Voice Onset Time (VOT) of Iranian learners of English. In the realm of pronunciation instruction, VF is gaining recognition as a novel approach. The present research study involved an experimental group, and statistical analysis, along with visual data inspection, revealed consistent performance across the study's three phases. Surprisingly, the VF paradigm did not lead to significant changes in VOT. However, there were noticeable variations among participants regarding their average VOTs, with some showing improvement in English VOTs for the phonemes /p/, /t/, and /k/, while others did not. In the pretest, most participants exhibited English-like VOTs, suggesting a potential ceiling effect. Factors such as immersion in the target language and the prominence of English likely influenced participants' VOT scores before the study. The findings of this study indicate that visual feedback can serve as a different instructional approach for teaching pronunciation in the classroom. This approach is more student-centered, as it enables students to analyse visual representations of their own speech. The participants' VOT ratings may have been influenced by the frequency at which they were exposed to the target language prior to the study. Nevertheless, in order to obtain more reliable conclusions, it is important to carry out studies including a larger number of subjects with varying levels of exposure.

ARTICLE INFO
Article history:
Received: 03 April 2024
Received in revised form 29 May 2024
Accepted: 29 May 2024
Available online: Spring 2024

Keywords:
Efficacy, Iranian EFL learners, Pronunciation, Visual feedback, Voice Onset Time.

Amirjalili, F. (2024). Investigating the Impact of Visual Feedback on Voice Onset Time (VOT) among Iranian EFL Learners. *Journal of Foreign Language Research*, 14 (1), 115-135. http//doi.org/ 10.22059/jflr.2024.374042.1115.

Publisher: The University of Tehran Press.
DOI: http//doi.org/ 10.22059/jflr.2024.374042.1115.

* Forough Amirjalili is an assistant professor in the Department of English Language and Literature at Yazd University, Yazd, Iran. She has received her Master's and Ph.D. degrees in Teaching English as a Foreign Language (TEFL) and holds a Bachelor's degree in English Language and Literature. Her main research interests include language teaching, second language acquisition and EFL teaching and learning.

1. Introduction

The importance of pronunciation in the second language classroom has been gaining recognition, contrary to its historical neglect. A growing body of literature, exemplified by the work of Derwing and Munro (2005), underscores the positive impact of pronunciation training on students' language production. Studies have shown that mispronunciations not only influence comprehensibility, but also affect intelligibility and perceived accentedness (Munro & Derwing, 2015; Derwing & Munro,2009, 2005). Recognizing these effects, classroom pronunciation training has been suggested as a valuable strategy to address these aspects of language learning.

Historically, the audiolingual method, as discussed by Celce-Murcia (2001) and Hauptman (1971), emphasized the importance of pronunciation as a key element in language courses. Initial endeavors involved the integration of pronunciation exercises such as repetition and drills (Celce-Murcia et al., 1997), contrasting similar sounds (Tuan, 2010; Levis & Cortes, 2008), and the utilization of the International Phonetic Alphabet (IPA) (Smith, 2000; Trask, 1996).

The recent advancements in technology have provided chances to increase feedback in language instruction by combining auditory and visual aspects (Bliss et al., 2018). Visual Feedback (VF) is a newly implemented tool for instructing pronunciation. Learners are able to visually observe their speech and then compare their speech to that of native speakers (Olson, in press). VF has demonstrated effectiveness in teaching both suprasegmental and segmental features, resulting in promising results in addressing voice onset time (VOT), particularly for voiceless stop consonants (/p, t, k/). The development of VOT presents a difficulty for Persian-speaking learners of English as a second language (L2) due to tiny yet meaningful disparities between the two languages.

Prior studies have extensively examined the use of VF to reduce VOT, as demonstrated by Offerman and Olson (2016) and Olson (2021, 2014). However, the current study aims to fill a research gap by exploring whether similar outcomes may be accomplished in prolonging VOT. Significantly, the study focuses on individuals who are native speakers of Persian (L1) and are learning English as a second language (L2). The study examines the potential problems in understanding caused by the differences between Persian and English in the way voiceless and voiced stops are spoken. For example, if the word "peach" is mispronounced with a Persian-like Voice Onset Time (VOT), native English speakers may perceive it as "beach." This study aims to investigate the efficacy of visual feedback (VF) in extending the voice onset time (VOT) of L1 Persian-speaking learners of L2 English. It provides useful insights into the use of VF to tackle specific pronunciation difficulties in cross-linguistic situations.

2.Literature Review

2.1. Voice Onset Time in English and Persian

Voice Onset Time (VOT) is a significant acoustic characteristic in phonetics that

measures the duration between the release of a stop consonant and the beginning of vocal cord vibration during speech production. It plays a significant role in distinguishing voiced and voiceless stops and contributes to the overall intelligibility and accentedness of an individual's pronunciation (Valipour, & Jamalzad, 2022). Persian and English exhibit distinct VOT norms for voiceless stops, with subtle but significant differences. Persian typically features shorter VOT values, while English has longer VOT values. This cross-linguistic variation can pose challenges for L1 Persian-speaking learners of L2 English, particularly in distinguishing and producing accurate VOT patterns. L1 Persian - L2 English learners commonly encounter difficulties related to VOT, given the inherent differences in the phonetic characteristics of Persian and English. Mispronunciations may lead to challenges in intelligibility, making the accurate production of VOT a crucial aspect of successful English pronunciation for this learner group (Jahani, & Korn, 2009). While traditional methods such as drills, repetition, and the use of the International Phonetic Alphabet (IPA) have been employed in pronunciation instruction, recent advancements in technology, particularly Visual Feedback (VF) through tools like Praat, have shown promise in enhancing VOT production. VF offers learners a graphical depiction of their spoken language, allowing for a more comprehensive understanding and analysis of VOT patterns (Valipour, & Jamalzad, 2022).

2.2 Pronunciation: Accentedness, Intelligibility, & Comprehensibility

Pronunciation refers to the combination of individual sounds and the overall patterns of speech in a specific language, as produced by individuals or groups of speakers (Trask, 1996; Pennington & Richards, 1986). According to Yates and Zielinski (2009), the primary objective of mastering pronunciation is to communicate meaning to others in an efficient manner. The principles of intelligibility, comprehensibility, and accentedness are of great importance in this particular setting (Munro & Derwing, 2015; Derwing & Munro, 1997, 2005).

According to Munro (2011), intelligibility refers to the degree to which a listener truly comprehends the pronunciation. Comprehensibility refers to the level of difficulty a listener experiences in understanding the pronunciation of a speaker. Accentedness pertains to the degree of perceptible distinctiveness in a speaker's pronunciation as perceived by others. Munro (2011) argues that the ability to be understood and understood completely are strongly connected to successful communication, and so deserve greater focus from educators. On the other hand, the importance of accentedness is diminished, as listeners have the ability to adjust to a broad spectrum of pronunciation patterns. Consequently, language learners can enhance intelligibility and comprehensibility without necessarily altering their accentedness (Derwing et al.,2009).

To summarize, pronunciation is acknowledged as a fundamental component in verbal communication across different languages (Macdonald, 2002). It encompasses not only the accurate articulation of phonetic elements but also

efficient interpersonal exchange (Barrera Pardo, 2004). Although pronunciation is important, it is generally given less attention in language teaching and is sometimes referred to as the "Cinderella" of language learning (Kelly, 1969; Underhill, 2013). The following section will examine the significance of pronunciation in the classroom, examining several methods used by various writers to incorporate L2 pronunciation training back into language schools.

2.3. **Teaching L2 pronunciation**

This section provides an overview of the existing literature on teaching pronunciation in the second language (L2) classroom. In Section 2.3.1, the text explores the significance of pronunciation in various teaching methodologies. Section 2.3.2 examines the various methods that teachers and researchers have used to incorporate pronunciation into the classroom setting.

2.3.1. ***L2 pronunciation instruction approaches***

The significance of pronunciation in language training has differed among various theoretical approaches (Levis, 2005). Several language teaching approaches, such as the Reading Method, the Grammar Translation Method, and Communicative Language Teaching (CLT), have largely neglected the incorporation of a substantial pronunciation element. On the other hand, both Both the Situational Approach and the Audiolingual Approach have constantly regarded pronunciation as an essential element.

The GTM focuses on the teaching of grammar. This methodology became increasingly popular in foreign language classes from the 1850s to the 1950s (Celce-Murcia, 1991). An integral part of the learning process entails actively attending to the teacher's explanations and participating in tasks that include completing missing information and translating between the learner's native language (L1) and their second language (L2). This traditional approach is still widely used by numerous educational institutions worldwide (Chang, 2011). Furthermore, the reading approach assigned a subordinate importance to pronunciation. This approach focused exclusively on teaching vocabulary and grammar that were relevant and commonly encountered in reading comprehension (Celce-Murcia, 2001). The Communicative Language Teaching (CLT) approach, developed by scholars such as Halliday, Hymes, Widdowson, and Wilkins in the late 1970s, was implemented due to concerns among educators and linguists that students were not acquiring language skills that would enable them to effectively communicate with individuals from different cultural backgrounds. This approach aimed to address the need for realistic and meaningful language use in intercultural interactions (Widdowson, 1972; Berns, 1983). According to Levis and Sonsaat (2019), there was a decrease in emphasis on teaching pronunciation as Communicative Language Teaching (CLT) gained popularity. The primary emphasis of this approach was on exercises that highlighted semantic conceptions and social functions, although it did provide students with the opportunity to engage in some communicative pronunciation tasks (Celce-Murcia, 1991).

Instructors who advocated for Communicative Language Teaching (CLT) frequently omitted the inclusion of pronunciation in their curricula, as the instruction of specific prosody and sounds did not align well with their preferred teaching approach (Foote et al., 2016).

In contrast, certain alternative methods have incorporated pronunciation as a fundamental aspect of their curriculum. Celce-Murcia (2001) explains that the audiolingual technique emerged in response to the reading approach, as described in her review. During the period from the 1940s to the 1960s, this approach was prevalent in the United States. It focused on the memorizing and imitation of language patterns in order to cultivate language proficiency in students. Furthermore, the teaching of grammar followed an inductive approach, whereas pronunciation received early emphasis. The situational method, which was widely acknowledged, incorporated pronunciation as a key element. Originating in Great Britain, this strategy evolved as a reaction to the dominant reading method and became more popular from the 1940s to the 1960s. Unlike the audiolingual method, which disregarded meaning and context in L2 learning, the situational approach helped students engage in authentic circumstances to experiment and enhance their oral talents (Hauptman, 1971; Celce-Murcia, 2001).

Although pronunciation is crucial, it is often overlooked in language instruction and learning, particularly in comparison to other dominant abilities like reading, listening, speaking, and writing (Underhill, 2013; Szyszka, 2017; Sweeting, 2021; Levis, 2005). Although certain teachers may be receptive to incorporating pronunciation activities into their curriculum, a significant number of them feel inadequately equipped to implement pronunciation approaches in their listening and speaking exercises (Sturm et al., 2019; Levis & Grant, 2003). Hence, it is widely acknowledged that teachers must possess expertise in pronunciation techniques in order to effectively apply them in their classrooms. By doing this, students can attain a degree of expertise in pronunciation that enables them to be easily understood by native speakers (Celce-Murcia, 2001). This underscores the importance of elevating the status of pronunciation within language instruction and emphasizing its integration into the broader language learning process.

2.3.1 Classroom pronunciation implementation

Teachers have endeavored to include pronunciation into their classrooms. Examples are drills and repetition, explanations of the IPA, and exercises with minimal pairs. Trask (1996) presented the International Phonetic Alphabet (IPA) as a means of instructing pronunciation. The International Phonetic Alphabet (IPA) was first developed as a uniform script for the analysis of phonetics and its use in practice (Smith, 2000). The International Phonetic Alphabet (IPA) is used in pronunciation education to provide students with a reference for accurate pronunciation. It enables students to consult dictionaries (Smith, 2000) and promotes learner autonomy (Mompean & Fouz-González, 2021).

Offerman (2016) highlighted IPA as a possible tool for instructing voiceless stop consonants in a study comparing different pronunciation strategies. The participants in this study showed progress in both activities that required more control and those that required less control, but they did not achieve a level of pronunciation comparable to that of native speakers. Nevertheless, the participants achieved noteworthy progress, as evidenced by participant surveys that showed heightened self-assurance in producing the sounds /p, t, k/.

Prior research has employed minimum pairs to examine pronunciation disparities between an individual's first language (L1) and their second language (L2). Minimal pairs consist of two words that differ by only one sound (Levis & Cortes, 2008; Brown, 1995). It is commonly believed that mispronouncing these particular words have the potential to cause misinterpretations (Levis & Cortes, 2008). However, Brown (1995) argues against the use of minimal pairs as the primary method for teaching pronunciation in the classroom. According to him, there are only a few minimal pairs that can be effectively taught, and activities focused on minimal pairs generally do not prioritize communication. According to the author, it is important to avoid giving too much importance to some characteristics of pronunciation, such as voice quality, intonation, rhythm, and stress, while neglecting other factors. This statement can be found on page 174.

Mompean and Fouz-González (2021) argue that phonetic symbols are especially effective in illustrating important characteristics of L2 pronunciation that would otherwise be challenging to depict. Essentially, the main objective of individuals acquiring a second language is to comprehend and effectively communicate with native speakers or have a high level of proficiency in that language. To accomplish this objective, one must possess not just a solid understanding of grammar but also essential proficiency in pronunciation (Tlazalo Tejeda & Basurto Santos, 2014). Although traditional approaches continue to be used, it is important to highlight that technology-enhanced pronunciation methods have become increasingly significant in recent years (Olson, 2014; Okuno & Hardison, 2016; Binasfour et al., 2017). The incorporation of technology highlights the evolving landscape of pronunciation instruction, offering innovative avenues for effective language learning.

2.4. **L2 classroom technology**

In recent decades, the widespread accessibility of computers worldwide has transformed the educational landscape, enabling language teachers to leverage technology for enhancing students' pronunciation skills (Eskenazi, 1999). Computer-Assisted Pronunciation Training (CAPT), first designed for speech pathologists aiding individuals with language difficulties, has been utilised in the field of pronunciation instruction (Pokrivcakova, 2014).

Historically, Visi-Pitch, a software developed by Kay Elemetrics in 1986, revolutionized the teaching of pronunciation in the late 1970s. It visually and analytically demonstrated how native speakers articulate

words and phrases in diverse languages. This groundbreaking technology was referenced by Albertson in 1982, Anderson-Hsieh in 1992, and Weltens & De Bot in 1984. This methodology, which combines the analysis of basic frequency and spectrograms, enables students to not only audibly perceive but also visually examine the way native speakers produce sounds, hence augmenting the importance of pronouncing acquisition (Okuno & Hardison, 2016).

The development of CAPT has brought forth new possibilities for providing feedback by integrating auditory and visual modes (Bliss et al., 2018). Visual Feedback (VF) in pronunciation education involves learners visually analyzing their own speech and comparing it to native speakers' speech (Olson, in press). Visual feedback (VF) can be categorized into two distinct groups: direct and indirect. Direct VF provides an immediate visual representation of the movements involved in articulation, whereas indirect VF presents audio information that allows learners to get a better understanding of their own articulation. (Olson, 2021; McCrocklin, 2012; Kartushina et al., 2015; Bliss et al., 2018).

A Visual Feedback Paradigm (VFP) typically comprises three elements: initial recording, visual and auditory analysis, and practice and re-recording. This system empowers learners to generate, analyze, and compare their recordings with those made by individuals who speak the language natively, promoting recognition of areas that need enhancement. Under the influence of the noticing hypothesis (Schmidt, 1994), the VFP emphasizes learners' awareness and attention to linguistic features in the second language, promoting explicit teaching for effective learning outcomes (Leow, 2018). The visual feedback paradigm is highly effective because it improves learners' self-awareness of their second language (L2) performances. This enables them to compare, correct, and imitate the linguistic qualities of native speakers across different aspects of language.

Praat (Boersma & Weenink, 2022) is a commonly used software package for delivering Visual Feedback (VF) in pronunciation education. Praat is widely recognized for its ability to allow students to record their own speech and obtain visual representations of their speech output in the form of waveforms, intonation contours, and spectrograms. It is worth mentioning that acquiring the skill to use Praat is regarded as a time-effective procedure (Olson, 2014).

The use of Praat as a methodological instrument has produced positive outcomes in the instruction of pronunciation. For example, Okuno and Hardison (2016) utilized visual feedback using Praat to teach students how to perceive vowel duration in Japanese. The study conducted a comparison among three groups: the auditory-only (A-only) group, the auditory-visual (AV) group, and the group that received no instruction. The findings indicated that both the groups exposed to audiovisual (AV) stimuli and audio-only (A-only) stimuli showed improvement in their ability to perceive the duration of Japanese vowels. Notably, the AV group had the highest level of improvement.

Highlighting the wider range of

possibilities that technology, namely (VF), offers in language training, it presents various opportunities, concepts, and approaches to enhance the pedagogical process in a language classroom (Nicolle & Lou, 2008). Utilizing technology in conjunction with the communicative approach aids learners in overcoming pronunciation challenges in their second language (L2), hence improving their ability to communicate effectively (Okuno & Hardison, 2016). Praat, among other technological tools, exemplifies how incorporating VF can contribute significantly to the improvement of pronunciation skills and, consequently, overall language proficiency.

The present study seeks to investigate the following research questions:

1. Does the provision of visual feedback enhance the generation of voice onset time (VOT) for intermediate-level L1 Persian – L2 English learners?

.I

2. Given that the training only targets isolated words, does the enhancement in VOT output for isolated words extend to connected speech?

3.Will the Iranian EFL participants retain the impacts of visual feedback instruction on voice onset time production after the study?

3.Methodology

3.1. Participants

The study included a homogeneous sample consisting of 24 undergraduate students from a prestigious university in Yazd, Iran. The university, a public institution, utilizes a national standardized test for student admissions. The participants, majoring in language and literature, are being prepared for future roles as foreign language and literature teachers, particularly in English. The major's curriculum is structured to progressively develop language proficiency and teaching skills. The language and literature major's curriculum unfolds in four initial semesters, primarily focusing on grammar and vocabulary components to ensure language proficiency. In the next semesters, the emphasis is placed on developing skills in oral communication, written expression, and teaching methods. The speaking component of the curriculum includes a specialized course on English phonetics, which is the only course that specifically focuses on pronunciation. In the later semesters, the curriculum explores subjects related to history and culture. Crucially, participants had not received explicit pronunciation instruction before this study, making them a particularly relevant group to investigate the impact of visual feedback on voice onset time (VOT) in voiceless stops. Participants completed the Bilingual Language Profile (Birdsong et al., 2012), a self-report instrument designed to assess language dominance in Persian and English. The tool covers participants' language use, language history, attitudes, and proficiency. The objective of the BLP was to create a comprehensive bilingual language profile by analyzing the answers provided by the participants. Table 1 presents the results of the Bilingual Language Profile, categorizing participants based on their language dominance in Persian and English. The table includes means and standard deviations (SDs) for participants' responses in various categories.

Table 1 Experimental group BLP subcomponent unweighted results

Component	Scale	English M (SD)	Persian M (SD)
History of Language	0–130	19.3 (3.9)	96.7 (4.8)
Language Usage	0–60	8.5 (1.4)	33.2 (2.3)
Language skill	0–29	14.9 (1.3)	20.6 (.5)
Attitudes towards language	0–29	15.2 (1.9)	20.8 (1.5)

All participants included in the study were native Persian speakers who commenced learning English at an average age of twelve (SD = 4.1). Their language learning trajectory encompassed eight years of English classes on average, extending from primary school through university (SD = 4.3). Participants reported variations in their daily English usage, with a predominant use of English at the university (35%), contrasting with minimal use within family contexts (4%) and moderate usage with friends (25%). The data indicates a lower frequency of English usage outside the university environment. After completing the Bilingual Language Profile (BLP) by Birdsong et al. (2012), participants engaged in tasks as part of their regular coursework. These tasks, integrated into their required coursework, were a mandatory component of their academic responsibilities. It is important to mention that the participants did not get any academic recognition for either successfully or unsuccessfully completing these activities. Participation in the present study, which required allowing access to their coursework for research reasons, was completely voluntary for the participants. Their decision to participate or not had no bearing on their academic standing or performance. The voluntary nature of participation underscores the ethical considerations of informed consent and autonomy for the participants.

3.2. Materials

3.2.1. Isolated words

The experimental design incorporated an Isolated Words Task, consisting of 35 words produced in isolation. These tokens were consistent across recording sessions, including the pretest, posttest, and delayed posttest. Every session included 12 tokens for each of the word-initial stops: /p/, /t/, and /k/. The words had two syllables and stress on the first syllable. Each word started with a stop consonant followed by one of the English vowels (/a, e, i, o, u/). This ensured an equal distribution of stimuli with two examples for each combination of consonant and vowel.

Table 2 Isolated-Tokens Example

	/p/			/t/			/k/
Token	IPA		Token	IPA		Token	IPA
Peanut	/ˈpiː.nʌt/		Tuna	/ˈtuː.nə/		Ketchup	/ˈketʃ.ʌp/
Porridge	/ˈpɔːr.ɪdʒ/		Target	/ˈtɑːr.gɪt/		Corner	/ˈkɔːr.nɚ/
Pepper	/ˈpep.ɚ/		Teacher	/ˈtiː.tʃɚ/		Calmness	/ˈkɑːm.nəs/

3.2.2. Words within Utterances

Task 2 encompassed the "Words within Utterances" activity, in which participants documented a grand total of ninety five utterances. The objective of this challenge was to create a situation where short speech segments included the specific target tokens (Elliott, 1997). A total of thirty-five distinct stimuli were employed in each of the three recording sessions, including the pretest, posttest, and delayed posttest. The phonetic environment examined the equitable distribution of vowels that come after the target stop consonant (/a, e, i, o, u/). This was done by ensuring that there were two distinct instances of each consonant-vowel combination across the three sessions. Table 4 presents a representative example of the stimuli employed in the Words within Utterances Task.

Table 3 Word-in-utterance example.

Stop	Example 1	Example 2
/p/	The day is **p**eaceful.	Older **p**eople typically have increased sleep duration.
/t/	I patronised a Japanese **t**eahouse.	I replace my **t**oothbrush on a monthly basis.
/k/	My aspiration is to become a **ch**emist in the future.	I require a **c**orkscrew to open this bottle of wine.

3.3. Data Collection and Analysis Procedures

The study included three Visual Feedback (VF) exercises, each consisting of a recording, an in-class comparison analysis, and a subsequent rerecording. Participants employed Praat (Boersma & Weenink, 2022) for both the initial recordings and the subsequent re-recording tasks. Visual feedback (VF) was only given to the experimental group during the in-class analysis of the VF activities.

Each VF activity followed a consistent structure:

1. **Recording:** Participants captured recordings of themselves articulating individual words as well as words within longer utterances.

2. **In-Class Comparative Analysis:** The group participated in a classroom analysis where they were given visual feedback for the three phonemes (/p/, /t/, /k/).

3. **Rerecording:** After the analysis conducted during the lesson, the participants proceeded to record themselves again, pronouncing various individual words as well as words within complete sentences. They had the freedom to record multiple times till they were satisfied with their productions.

3.3.1. Activity 1

This stage consisted of a thorough series of elements, which included a recording, analysis in the classroom, and a subsequent recording. The initial recording in VF Activity 1 included both individual words and words inside complete sentences, which were aligned with the target words in the pretest. In VF Activity 1, the participants captured their pronunciation of the target words in both isolated form and inside

sentences during the initial recording. This recording set the foundation for subsequent analyses and comparisons. During the in-class analysis, participants engaged with a handout containing reflective questions designed to guide them in evaluating their productions. The handout featured group-oriented questions related to the images of the word "Parlor" those participants had recorded. The questions covered aspects such as marking sound boundaries, visual characteristics of the "p" sound, and the length comparison between "p" and "a."

Examine the images of the term *Parlour* that you documented in groups of 2-3, and respond to the following questions:

What was your method for determining the limits of each individual sound?

What are the visual attributes of your "p"? Is the length of it brief or extensive?

Is the duration of your "p" sound longer or shorter compared to the duration of your "a" sound?

Following this group activity, participants were presented with native speaker (NS) productions for comparative analysis. This entailed comparing student productions with NS samples, including characteristics such as larger/smaller or darker/lighter to enhance comprehension of waveforms and spectrograms. Upon finishing the handout and the in-class analysis, participants were assigned the task of re-recording a list of additional isolated words. They were instructed to send these recordings to a specified email address and take screenshots of six of the words for presentation in the next session.

3.3.2. Activity 2

This stage closely resembled VF Activity 1 in terms of organization, but the main difference was that all activities were carried out using non-target tokens. Despite the shift to non-target tokens, the inclusion of words beginning with /p/, /t/, /k/, and a balanced distribution of /a/, /e/, /i/, /o/, /u/ was maintained. The aim was to provide participants with a varied set of linguistic stimuli for analysis and improvement. Participants participated in a recording session where they enunciated a collection of individual words as well as terms inside longer phrases. Participants in VF Activity 1 captured screenshots of six individual tokens and recorded the precise borders of each token in a Word document. The ensuing in-class analysis consisted of participants responding to the identical set of reflective questions offered in VF Activity 1. This facilitated a consistent framework for self-reflection and comparison with native speaker (NS) productions. Following the analysis conducted during the class session, participants were given instructions to return home and record themselves again, with the objective of observing any possible enhancements in their performance. Subsequently, they were obligated to capture screenshots of six words from the list, delineate their limits, and exhibit them in the following session.

3.3.3. *Activity 3*

Participants initiated the recording session by pronouncing a list of words both in isolation and within utterances. This recording set the foundation for the subsequent analysis and evaluation. The in-class analysis involved participants being

JOURNAL OF FOREIGN LANGUAGE RESEARCH, Volume 14, Number 1, Spring 2024, Page 115 to 135

presented with waveforms and spectrograms of the same words pronounced by a native speaker of English. This visual aid aimed to provide a clear comparison between participants' productions and those of a proficient English speaker. Participants completed a handout containing questions designed to guide them in comparing their productions to those of the NS. This reflective process encouraged participants to critically assess their own pronunciation based on visual representations. After the analysis conducted during the class, the participants were assigned the task of recording a list of controlled words both individually and inside sentences. Subsequently, it was necessary to transmit the recorded files to the designated email address. The last recording in VF Activity 3 matched the target words in the posttest. The posttest phase enabled researchers to compare participants' articulation of voiceless plosives (/p/, /t/, /k/) from the pretest to this stage.

3.3.4. Final Activity

The component of the trial, VF Activity 4, included a delayed posttest that was conducted one month following the rerecording of VF Activity 3. The researcher reached out to the participants to invite them to take part in VF Activity 4, which included a posttest that was conducted at a later time. During this phase, the participants independently recorded their pronunciation of the same words in isolation, which were also utilized in the pretest and posttest. Furthermore, a fresh compilation of words used in spoken expressions was introduced. The purpose of the delayed posttest was to assess the participants' ability to maintain the improvements achieved in the pronunciation of the English voiceless stops (/p/, /t/, /k/) after completing the three VF activities. Participants were allotted a one-week period to finish the task within the convenience of their own residences.

4.Results

4.1. Isolated words

During the examination of individual words that contain English voiceless stops (/p/, /t/, and /k/), a total of 2208 instances out of a potential 2423 [pre-test, post-test, delayed post-test] were considered. The deletion of 172 tokens was due to several variables. These considerations consist of three categories: instances of noise that rendered the measurement of VOT difficult (n = 41), cases where participants failed to produce the target tokens (n = 109), and outliers defined as VOTs that deviated more than ±3 standard deviations from the mean (n = 12).

When examining the phoneme results, a comparable pattern becomes evident. A total of 698 words were produced in isolation for the phoneme /p/. The average Voice Onset Time (VOT) for these words was 29.7 milliseconds, with a standard deviation of 24.6 milliseconds. Figure 1 demonstrates that the average time in the pretest was 34.7 milliseconds, with a standard deviation of 27.3 milliseconds. The average time in the posttest was 27.6 milliseconds, with a standard deviation of 23.4 milliseconds. Ultimately, in the postponed posttest, their mean was 31.6 ms (standard deviation = 19.1 ms). A total of 695 words were created in isolation for the /t/ sound, with an average

Voice Onset Time (VOT) of 45.3 milliseconds and a standard deviation of 26 milliseconds. Figure 2 shows that the average score in the pretest was 46.6 ms with a standard deviation of 25.7 ms. The participants' average response time in the posttest was 45.3 milliseconds, with a standard deviation of 27.5 milliseconds. The average time in the delayed posttest was 43.2 milliseconds, with a standard deviation of 25.1 milliseconds. A total of 718 words were produced in isolation for the sound /k/. The average Voice Onset Time (VOT) for these words was 74.8 milliseconds, with a standard deviation of 26.7 milliseconds. Figure 3 displays the average response time of the participants in the pretest, which was 81 ms with a standard deviation of 27.4 ms. The participants' average response time in the posttest was 73.2 milliseconds, with a standard deviation of 23.1 milliseconds. Ultimately, the delayed posttests yielded an average of 76.6 milliseconds, with a standard deviation of 26.3 milliseconds.

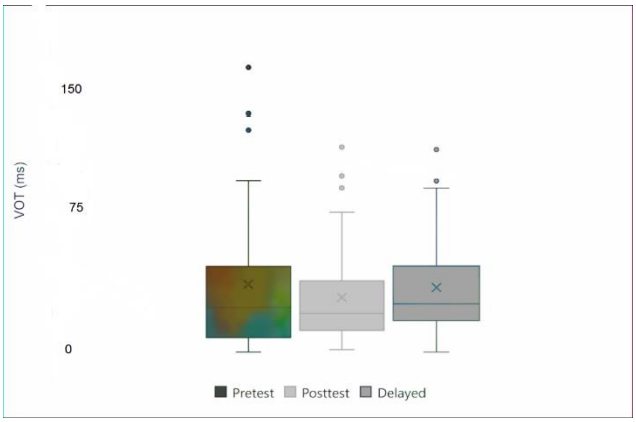

Figure 1. Words in isolation: /p/

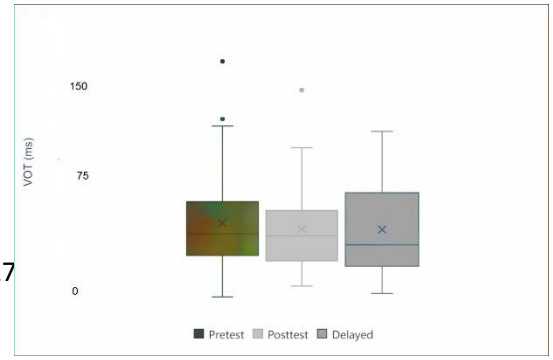

Figure 2. Words in isolation: /t/

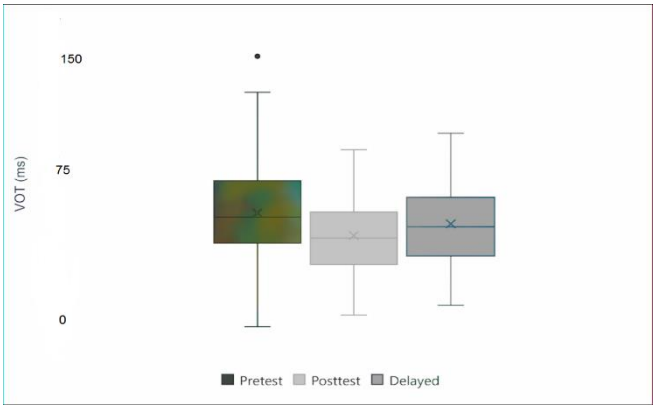

Figure 3. Words in isolation: /k/

LME4 was used to analyze data with R Statistical Software (R Development Core Team, 2008) (Bates, Maechler, Bolker, & Walker, 2015). Linear Mixed Effects Model results showed no significant variations in Voice Onset Time (VOT) between pretest and posttest ($\beta = 1.203$, $t = 0.503$) or delayed posttest ($\beta = 2.401$, $t = .886$). Visual evaluation of the data and statistical analysis show that the experimental group performed similarly in all three phases of the study. VOT did not change with the visual feedback paradigm.

4.2. Utterance words

A total of 2208 tokens out of a potential 2423 [pre-test, post-test, delayed post-test] were analyzed for words in utterances containing English voiceless stops /p/, /t/, and /k/.

1. /p/: A total of 6۸۳ words were spoken, with an average Voice Onset Time (VOT) of 29.4 milliseconds (standard deviation = 23.4 milliseconds). The group demonstrated an average of 33.1 milliseconds (standard deviation = 26.9 milliseconds) in the pretest. In the posttest, the mean value rose to 27.4 milliseconds with a standard deviation of 22.5

milliseconds. In the delayed posttest, the mean value reached 30.4 milliseconds with a standard deviation of 21.3 milliseconds (Figure 4).

2. **/t/:** A total of 703 words were said, with an average Voice Onset Time (VOT) of 51.2 milliseconds (standard deviation = 25.5 milliseconds). The participants started the pretest with an average of 53.3 milliseconds, with a standard deviation of 28.7 milliseconds. The average in the posttest reduced to 50.1 ms with a standard deviation of 21.1 ms. In the posttest that was conducted later than expected, the value increased to 54.2 milliseconds with a standard deviation of 27.4 milliseconds, as shown in Figure 5.

3. **/k/:** A total of 691 words were spoken, with an average Voice Onset Time (VOT) of 76.2 milliseconds, and a standard deviation of 23.5 milliseconds. The average in the pretest was 73.6 milliseconds, with a standard deviation of 26 milliseconds. In the posttest, there was a rise to 80.4 ms (standard deviation = 25 ms), and in the delayed posttest, the mean was 74.1 ms (standard deviation = 19.4 ms) (Figure 6).

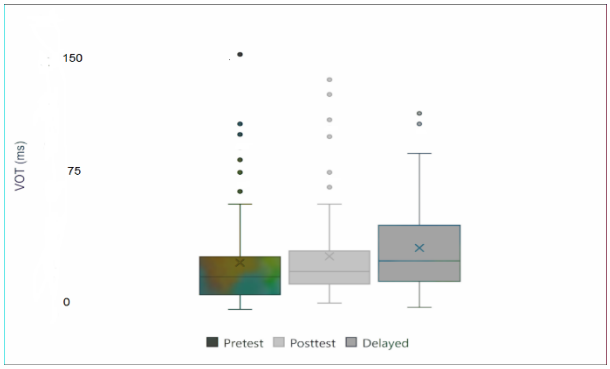

Figure 4. Words in utterances /p/

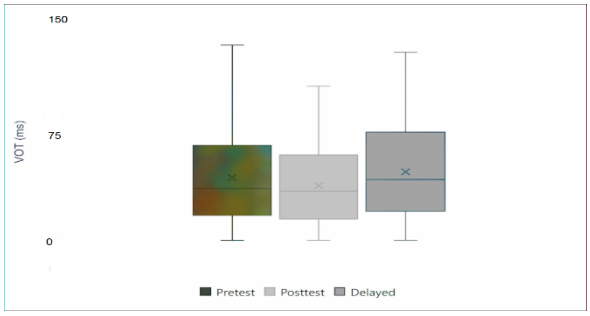

Figure 5. Words in utterances /t/

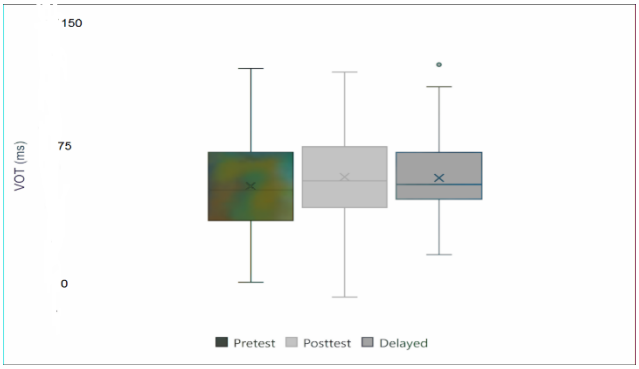

Figure 6. Words in utterances /k/

R and the LME4 package (Bates, Maechler, Bolker, & Walker, 2015) were used for statistical analysis. A Linear Mixed Effects Model with Time as a fixed factor and Participants and Phoneme as random factors did not reveal significant differences in VOT between pretest (intercept) and posttest (β = 3.207, t = 1.408) and delayed posttest (β = 2.103, t = 1.223). The experimental group improved Voice Onset Time (VOT) for English voiceless stops /p/, /t/, and /k/, but not significantly. This implies that visual feedback did not improve connected speech with time.

4.3. Individual outcomes

The statistical analysis did not reveal any substantial alteration in VOT for the full set of participants as a result of the VF paradigm, but there was a discernible degree of individual variation in the findings. The initial group included individuals who consistently exhibited an increase in their reaction time for the three phonemes /p/, /t/, and /k/ during all three phases of the trial. In contrast, the second group comprises the patients who did not exhibit a rise in their Voice Onset Times for the same phonemes across the three phases of the study.

Participants 9, 18, 22, and 24 exhibited

a rise in their Voice Onset Times (VOTs) from the initial assessment to the subsequent assessment while analyzing individual words. Participants 3, 5, 6, 7, 8, 10, 13, 20, and 21 experienced an increase in their Voice Onset Time (VOT) averages from the pretest to the posttest. Nevertheless, these mean values declined from the posttest to the delayed posttest. On the other hand, participants 1, 3, 5, 12, 13, 18, and 20 experienced a decrease in their Voice Onset Times (VOTs) between the first test and the delayed follow-up test. Participants 2, 4, 15, 16, 17, and 24 demonstrated a decrease in their Voice Onset Time (VOT) from the initial assessment to the final assessment. Nevertheless, the Voice Onset Times (VOTs) exhibited a rise from the posttest to the delayed posttest. Participants 1, 13, 15, and 20 exhibited a rise in their Voice Onset Time (VOT) for words in their utterances from the initial test to the postponed follow-up test. Participants 3, 6, 7, 11, 14, 16, 18, 21, 23, and 24 had enhanced general Voice Onset Times (VOTs) from the pretest to the posttest. However, the Voice Onset Times (VOTs) showed a decrease from the posttest to the delayed posttest. In contrast, Participants 5, 9, and 12 demonstrated a reduction in their mean Voice Onset Times (VOTs) from the initial assessment to the subsequent delayed follow-up assessment. In contrast, individuals 2, 5, 11, 13, 16, and 24 exhibited a decrease in VOTs from the initial test to the follow-up exam. However, they experienced an increase in VOTs from the follow-up test to the delayed follow-up test.

Contrary to expectations, those who had a lower average Voice Onset Time (VOT) did not show the largest shift. From the data provided, it is evident that all individuals had varying degrees of shift, regardless of their initial VOT lengths for both isolated words and words in sentences. In addition, the participants began with a mean Voice Onset Time (VOT) of 51.7 ms (standard deviation = 13.6 ms) for individual words and a mean of 49.5 ms (standard deviation = 13.4 ms) for words in complete sentences. Upon initial inspection, these findings suggest that all the individuals' English Voice Onset Time (VOT) measurements were within the expected range. It is possible that the advantages experienced by one subset of individuals were offset by the other members of the group, leading to the perception that the influence of visual stimuli on them was underestimated.

5.Discussion

The selection of English voiceless stops as the primary subject of investigation was motivated by the significant disparity in Voice Onset Time (VOT) observed between English and Persian voiceless stops. English voiceless stops generally exhibit a VOT (Voice Onset Time) with a longer duration, ranging from 30 to 100 milliseconds. In contrast, Persian voiceless stops tend to have a shorter VOT, commonly ranging from 0 to 30 milliseconds. The participants were cognizant of this distinction and its possible influence on the clarity, understandability, and foreignness of L2 English speech.

For instance, When Persian-speaking learners of English as a second language (L2) pronounce the word "peach" with a Voice Onset Time (VOT) that resembles Persian, it may be perceived as "beach". Moreover,

unvoiced pauses can be readily distinguished in a spectrogram when compared to other phonemes. As a result, it was anticipated that the participants would have no difficulty in mentally visualizing VOT. The experimental findings indicate that the tested group exhibited comparable performance across all three phases of the inquiry. Additionally, the use of visual feedback did not cause any changes in VOT. On average, individuals initially had a Voice Onset Time (VOT) of 54.2 ms (standard deviation = 14.6 ms) for isolated words. This indicates that the majority of participants already had a tendency to create long VOTs during the pretest, suggesting a ceiling effect. Ceiling effects refer to participant scores that approach or reach the uppermost limit, as defined by Garin (2014). The causes for these effects can vary due to factors such as the relatively straightforward nature of the test and the existing high level of competency within the testing group. When there are ceiling effects, it is not possible to accurately assess or establish the full degree of individuals' talents (Wang et al., 2009; Uttl, 2005). For this study, the majority of the participants generated Voice Onset Times (VOTs) that resembled those of the English language right from the start of the trial. Therefore, there was no discernible progress to evaluate in the posttest and delayed posttest for individual terms.

Individual differences refer to the differences observed among participants in how a characteristic responds to an experiment. Research involves two types of variability: between-individual variability, which refers to changes among participants within the same research, and intra-individual variability, which refers to differences observed in the same participant at different time points during the study. This study primarily focused on the variability within people. We anticipated that participants would begin with a similar duration of Voice Onset Time (VOT) and show improvement over the course of the trial. Nevertheless, the subjects attained a point of maximum performance in the pretest. Therefore, there was no discernible enhancement to quantify; they were already generating Voice Onset Times (VOTs) that resembled those of English.

What we discovered was a situation characterized by differences among individuals, namely between participants who raised their Voice Onset Times (VOTs) and participants who lowered their VOTs. Among the 14 participants, only four managed to sustain their Voice Onset Times (VOTs) throughout the delayed posttest, while the remaining subjects had an increase in their VOTs. By contrast, 15 participants had a decrease in their Voice Onset Times on the posttest, and 8 of them consistently maintained their VOTs on the delayed posttest. over the posttest, fifteen subjects experienced an increase in their Voice Onset Time (VOT), while three individuals maintained consistent VOTs over the delayed posttest. In contrast, 8 participants reduced their initial Voice Onset Time (VOT) measures on the posttest, and two of them showed similar VOT outcomes on the delayed posttest. Based on this data, it is clear that the participants in this study had extended Voice Onset Times (VOTs) at the

beginning. Nevertheless, there was divergence among individuals, with certain individuals augmenting their Voice Onset Time (VOT) and others diminishing it. Consequently, the group's final outcomes exhibited a wide range of variations and were difficult to anticipate.

6. Conclusion

The presence of visual feedback did not impact the performance of the experimental group participants in producing the sounds /p/, /t/, and /k/ at all three stages of the study (pretest, posttest, delayed posttest). From the outset of the study, the participants exhibited Voice Onset Time (VOT) values that closely matched those suggested by Lisker and Abramson (1964) for English (30-100 ms), indicating a high level of performance with limited room for improvement. Due to the ceiling effects observed in participants' word production in isolation, it was not feasible to address the study inquiries on the generalization of these effects to words in utterances and the long-term retention of improvements following the experiment. Thus, it seemed that VOT was not pertinent for training among this particular group of participants. However, the utilization of visual feedback in the classroom facilitated a student-centered approach to teaching pronunciation. The tool allowed students to examine graphic depictions of their speech, such as waveforms and spectrograms, and make comparisons with those of individuals who are native speakers. This allowed them to identify and correct any errors. The participants' VOT ratings may have been influenced by the frequency at which they were exposed to the target language prior to the study. However, in order to obtain more reliable conclusions, it would be imperative to carry out studies including a larger number of subjects with varying levels of exposure. Moreover, the prominence of the target language in relation to Voice Onset Time (VOT) could have been another element that affected participants' knowledge of this attribute from the beginning of their learning journey.

The study's most prominent constraint is the participants' level of ability in pronouncing the target language. As previously mentioned, the participants already had a high level of pronunciation skill at the time of the experiment. This is why they were able to reach the maximum level of performance in the pretest. The participants' high proficiency in English can be attributed to factors such as their extensive exposure to the language as students in a language teaching program, the prominence of English in terms of Voice Onset Time (VOT) compared to Persian, or other factors that we may have overlooked. Considering this, the study might have been improved by included individuals who have had less exposure to English in a teaching context. Based on the study's findings and the aforementioned reasons, it seems that VOT is not applicable for instructing this particular population. Participants successfully generated the desired Voice Onset Times (VOTs) for all phonemes (/p, t, k/) in all study assessments. This indicates that the participants already possessed this characteristic prior to the experiment, implying that students should prioritize their time and effort on other features that are more difficult to acquire in

their second language. However, further research including a larger number of participants with varying degrees of skill is necessary to determine whether VOT is indeed irrelevant for teaching English to Persian learners at different levels. Additionally, this research should investigate the potential factors that may contribute to this observed pattern.

References

جمال زاد, محمد. (۱۴۰۱). بررسی ویژگی‌های صوت‌شناختی همخوان‌های سایشی در زبان روسی در بیان فارسی دانشجویان . ***پژوهشهای زبانشناختی در زبانهای خارجی***, *12*(3), 238-256. doi: 10.22059/jflr.2020.307078.740

Albertson, K. (1982). Teaching pronunciation with visual feedback. *IALLT Journal of Language Learning Technologies*, *17*(1), 18–33.

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, *20*(1), 51–62. https://doi.org/10.1016/0346-251X(92)90007-P

Barrera Pardo, D. (2004). Can pronunciation be taught?: A review of research and implications for teaching. *Revista Alicantina de Estudios Ingleses: RAEI*, *17*(17), 6–38. https://doi.org/10.14198/raei.2004.17.03

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berns, M. S. (1983). Functional Approaches to Language and Language Teaching: Another Look.

Binasfour, H., Setter, J. & Aslan, E. (2017). Enhancing L2 learners' perception and production of the Arabic emphatic sounds. In *Phonetics Teaching and Learning Conference*, 16–20.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual language profile: An easy-to- use instrument to assess bilingualism*. COERLL, University of Texas at Austin, 2013, from https://sites.la.utexas.edu/bilingual/

Bliss, H., Abel, J., & Gick, B. (2019). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation*, *4*(1), 129–153. https://doi.org/10.1075/jslp.00006.bli

Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer [Computer Program]. *Retrieved from* http://www.praat.org/.

Brown, J. D. (1995). The elements of language curriculum: A systematic approach to program development. Heinle & Heinle Publishers, 20 Park Plaza, Boston, MA 02116.

Chang, S. C. (2011). A contrastive study of grammar translation method and communicative approach in teaching English grammar. *English language teaching*, *4*(2), 13.

Celce-Murcia, M. (2001). Language teaching approaches: An overview. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language*, *2*(1), 3–10.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1997). *Teaching pronunciation a reference for teachers*

of English to speakers of other languages. Cambridge University Press.

Celce- Murcia, M. (1991). Grammar pedagogy in second and foreign language teaching. *TESOL quarterly*, *25*(3), 459-480.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397. https://doi.org/10.2307/3588486

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476–490.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company. https://doi.org/10.1075/lll.42

Elliott, A. R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, *80*(1), 95–108. https://doi.org/10.2307/345983

Eskenazi, M. (1999). Using a computer in foreign language pronunciation training: What advantages? *CALICO Journal*, *16*(3), 447–469. https://doi.org/10.1558/cj.v16i3.447-469

Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, *44*(2), 181-196.

Garin, O. (2014). Ceiling effect. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 631–633). https://doi.org/10.1007/978-94-007-0753-5_296

Halliday, M. A. K. (1973). Explorations in the functions of language.

Hauptman, P. C. (1971). A structural approach vs. A situational approach to foreign-language teaching. *Language Learning*, *21*(2), 235–244. https://doi.org/10.1017/S026144480800551X

Hymes, D. (1972). On communicative competence. *sociolinguistics*, *269293*, 269-293.

Jahani, C., & Korn, A. (2009). Balochi. In G. Windfuhr (ed.), *Iranian languages* (pp. 634-692). London, New York: Routledge

Kelly, L. G. (1969). *25 centuries of language teaching; An inquiry into the science, art, and development of language teaching methodology, 500 B.C.-1969*. Newbury House Publishers.

Leow, R. P. (2018). Noticing hypothesis. *The TESOL Encyclopedia of English Language Teaching*, *1*, 1-6. https://doi.org/10.1002/9781118784235.eelt0086.pub2

Levis, J., & Cortes, V. (2008). Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching. *Towards adaptive CALL: Natural language processing for diagnostic language*

assessment, 197208.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377. https://doi.org/10.2307/3588485

Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal*, *12*(2), 13–19. https://doi.org/10.1002/j.1949-3533.2003.tb00125.x

Levis, J. M., & Sonsaat, S. (2019). Quality teacher education for pronunciation teaching in L2 classrooms. In *Quality in TESOL and teacher education* (pp. 213-222). Routledge.

Li, H., Chen, Z., & Zhu, W. (2019). Variability: Human nature and its impact on measurement and statistical analysis. *Journal of Sport and Health Science*, *8*(6), 527–531. https://doi.org/10.1016/j.jshs.2019.06.002

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, *20*(3), 384–422.

Macdonald, S. (2002). Pronunciation-views and practices of reluctant teachers. *Prospect*, *17*(3), 3–18. https://doi.org/10.3316/aeipt.124108

McCrocklin, S. (2012). Effect of audio vs. video listening exercises on aural discrimination of vowels. *Pronunciation in Second Language Learning and Teaching Proceedings, 3*(1), 178-186.

Mompean, J. A., & Fouz-González, J. (2021). Phonetic symbols in contemporary pronunciation instruction. *RELC Journal*, *52*(1), 155–168. https://doi.org/10.1177/0033688220943431

Munro, M. J. (2011). Intelligibility: Buzzword or buzzworthy? In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference* (pp. 7–16). Ames, IA: Iowa State University.

Nicolle, P. S., & Lou, Y. (2008). Technology adoption into teaching and learning by mainstream university faculty: A mixed methodology study revealing the "how, when, why, and why not." *Journal of Educational Computing Research*, *39*(3), 235–265. https://doi.org/10.2190/EC.39.3.c

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second language segmental production: The generalizability of pronunciation gains. *System*, *59*, 45–60. https://doi.org/10.1016/j.system.2016.03.003

Okuno, T., & Hardison, D. M. (2016). Perception–production link in L2 Japanese vowel duration: Training with technology. *Language Learning & Technology*, *20*(2), 61–80.

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second–language pronunciation instruction. *Hispania*, *97*(1), 47–68. https://doi.org/10.1353/hpn.2014.0030

Olson, D. J., & Offerman, H. M. (2021).

Maximizing the effect of visual feedback for pronunciation instruction: A comparative analysis of three approaches. *Journal of Second Language Pronunciation*, *7*(1), 89–115. https://doi.org/10.1075/jslp.20005.ols

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly*, *20*(2), 207–225. https://doi.org/10.2307/3586541

Pokrivcakova, S. (2014). CALL and teaching pronunciation. *CALL and Foreign Language Education*, 29–37. https://doi.org/10.17846/CALL.2014.29–37

Schmidt, R. W. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 165–209). London: Academic Press.

Smith, C. L. (2000). Handbook of the international phonetic association: a guide to the use of the International Phonetic Alphabet. *Phonology*, *17*(2), 291–295. https://doi.org/10.1017/S0952675700003894

Sturm, J. L., Miyamoto, M., & Suzuki, N. (2019). Pronunciation in the L2 French classroom: Student and teacher attitudes. *The French Review*, *92*(3), 60–78. https://doi.org/10.1353/tfr.2019.0182

Sweeting, A. (2021). *From "Cinderella" to "Sleeping beauty": Taking pronunciation instruction in the CELTA course beyond "Listen and repeat"* [Doctoral dissertation, University of the Sunshine Coast, Queensland]. https://doi.org/10.25907/00040

Szyszka, M. (2017). *Pronunciation learning strategies and language anxiety: In search of an interplay*. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-50642-5

Tlazalo Tejeda, A. C., & Basurto Santos, N. M. (2014). Pronunciation instruction and students' practice to develop their confidence in EFL oral skills: La instrucción de la pronunciación y la práctica de los estudiantes para el desarrollo de la confianza en habilidades orales en inglés como lengua extranjera. *PROFILE Issues in Teachers' Professional Development*, *16*(2), 151–170.

Trask, R. L. (1996). *A dictionary of phonetics and phonology*. Routledge.

Tuan, L. T. (2010). Teaching English discrete sounds through minimal pairs. *Journal of Language Teaching and Research*, *1*(5), 540–561. https://doi.org/10.4304/jltr.1.5.540-561

Underhill, A. (2013). Cinderella, integration and the pronunciation turn. *Speak out*, *49*(4), 4–8.

Uttl, B. (2005). Measurement of individual differences: Lessons from memory assessment in research and clinical practice. *Psychological Science*, *16*(6), 460–467. https://doi.org/10.1111/j.0956-7976.2005.01557.x

Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2009). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research*, *43*(3), 476–496. https://doi.org/10.1080/00273170802285941

Weltens, B., & De Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language & Speech*, *27*(1), 79–88.

Widdowson, H. G. (1972). The teaching of English as communication. *ELT journal*, *27*(1), 15-19.

Wilkins, D. A. (1972). *Linguistics in language teaching* (Vol. 111). London: Edward Arnold.

Yates, L., & Zielinski, B. (2009). *Give it a go: Teaching pronunciation to adults*. Macquarie University.