



Gender differences in the acquisition of the VOT of English bilabial stops among Saudi ESL learners

Wasmiah Alsuheim^{a 1} 

^a Department of English Language & Translation, College of Arabic Language & Social Studies,
Qassim University, Buraidah, Saudi Arabia

APA Citation:

Alsuheim, W. (2022). Gender differences in the acquisition of the VOT of English bilabial stops among Saudi ESL learners. *Journal of Language and Linguistic Studies*, 18(Special Issue 1), 100-107.

Submission Date: 17/08/2021

Acceptance Date: 20/10/2021

Abstract

This investigation explored the voice onset time (VOT) values of word-initial bilabial stops /p/ and /b/ as realized by Arabic-speaking male and female Saudi ESL learners. It sought to identify the differences, if any, between these two groups with regard to this acoustic measure (i.e., VOT). The VOT measurements were then compared with those of native English speakers to test the contrastive analysis (CA) hypothesis. The VOTs were measured from the recordings of the 20 participants (10 females and 10 males). This study used the Praat speech analysis software to measure VOT values. The results revealed a statistically significant difference between male and female participants in their VOT values for /b/. For the VOT values of /p/, no statistical difference was found in the results. Further, an increase was detected in the male participants' VOTs for /b/ compared with those of native English speakers. The remainder of the results were relatively in line with the VOT values of native English speakers; namely, both males and females acquired the voiceless phoneme /p/ despite the lack of this phoneme in their native language (Arabic).

Keywords: voice onset time; Saudis; ESL; bilingualism; English bilabial stop

1. Introduction

Arabic and English are distinguished by their very distinct linguistic systems, which derive from their different language families. English is from the Indo-European family of languages (Al-Shujairi et al., 2015), while Arabic is a Semitic language (Nofal, 2011). In a second language (L2) learning context, L2 acquisition tends to be affected by learners' first language (L1) as well as their target language (TL). One of the key notions of how L2 acquisition operates is the contrastive analysis hypothesis (CA); namely, that differences between learners' L1 and their target L2 cause difficulties in acquiring the latter while similarities between them ease the learning process (Lado, 1976). This hypothesis has become the focus of many linguists investigating the cross-linguistic influences between learners' L1 and their target L2 (Richards & Schmidt, 2010).

In the context of the present study, Arabic and English differ in their voiced and voiceless sounds; while Arabic lacks the English voiceless bilabial /p/, it features a voiced /b/. This difference tends to

¹ Corresponding author.

E-mail address: wasmiah.s@hotmail.com

result in Arabic-speaking English as a second language (ESL) learners being able to phonate /b/ easily while having difficulties with /p/; specifically, “the tendency is for the native speaker /p/ sound to be replaced by a /b/ sound by late Arab bilinguals” (Askoul, 2017, p. 75).

Over the past few decades, many studies have investigated this voicing contrast by measuring the differences in speaker’s stop consonants using voice onset time (VOT) (e.g., Abdelaal, 2017; Askoul, 2017; Koenig, 2000; Korkmaz & Boyaci, 2018; Lisker & Abramson, 1964; Rae, 2018; Sweeting & Baken, 1982; Yu et al., 2015). Some studies focused on examining cross-language VOT values (e.g., Abdelaal, 2017; Askoul, 2017), while others investigated gender-based differences in VOT measurements (e.g., Korkmaz & Boyaci, 2018; Sweeting & Baken, 1982; Yu et al., 2015). To explore cross-language VOT values, Abdelaal (2017) analyzed the VOTs for English stops produced by ten L1 Arabic ESL learners. The researcher sought to examine the effect of the participants’ L1 on their production of /p, b, t, d, k, g/; and the extent to which Arab ESL learners distinguish between the pronunciation of minimal pairs (e.g., *bat* and *pat*). The participants were asked to phonate the words *pat*, *bat*, *tap*, *dap*, *cap*, and *gap* in the carrier phrase “Say ... again”. The findings showed no significant differences between the participants’ pronunciations of /d/ and /t/ or between /g/ and /k/. However, the findings revealed that the participants could distinguish between /p/ and /b/ in a relatively similar way to native English speakers (NESs); thus, the assumption that Arab ESL learners have difficulties pronouncing the phoneme /p/ is contradicted.

Similarly, Askoul (2017) studied the production of the bilabial stop /p/ by Arab ESL learners. The data were elicited from 30 L1 Arabic-speaking late bilinguals using three oral tasks: (i) a series of picture descriptions, (ii) timed picture descriptions, and (iii) word reading. In these tasks, a total of 12 target words were exploited in which the phoneme /p/ was placed initially in a CVC context. The analysis revealed that the production of the phoneme /p/ by the participants deviated from that of NESs. The findings showed an overall tendency for the participants to pronounce /p/ with a small VOT value resembling the VOT of /b/. Thus, Askoul’s (2017) study supports the assumption that L1 Arabic-speaking ESL learners face difficulties acquiring the phoneme /p/, which is contradicted by Abdelaal’s (2017) findings.

To explore gender differences in VOT values, Sweeting and Baken (1982) analyzed the English bilabial stops /p/ and /b/ produced by 30 adult participants divided into three age groups: 25–39 years old ($N = 10$), 65–75 ($N = 10$), and 75 and over ($N = 10$). Each of these groups was equally composed in terms of gender. The researchers measured the target phonemes using three stimulus words (*bead*, *pete*, *beat*) in a carrier phrase. The findings revealed that the differences in VOT values of /p/ and /b/ were not statistically significant among the groups. That is, no statistically significant age effect was detected, nor was there any impact of gender on the participants’ VOT values.

Korkmaz and Boyaci (2018) also investigated gender differences in the VOT values of six stop consonants /p, b, t, d, k, g/ produced by 20 Turkish participants. The target phonemes featured in 48 CVC word-initial stimulus words. All the data were analyzed using the MATLAB classification software. The findings revealed the existence of gender differences in the participants’ productions of the investigated phonemes. The findings highlighted that gender differences were most evident in the VOT values for the voiceless stops /p, t, k/. However, Korkmaz and Boyaci’s (2018) study failed to clarify the exact nature of these differences: the researchers did not state the mean VOT values for the investigated phonemes, nor did they discuss whether male or female participants produced higher VOT values.

Similarly, Yu et al. (2015) explored age and gender differences in the production of voiceless stops /p, t, k/ by analyzing VOT values produced by NESs. A total of two stimuli, /pa/ and /pataka/, were pronounced by 70 children (aged 4.1–18.4) and 12 adults (mean age = 27.1). In terms of the age effect

on the measured VOTs, the findings showed a negative correlation between mean VOT values and participant age. Additionally, the findings indicated that participants over 11 years old produced adult-like VOT patterns. In terms of gender differences in VOT values, the study found higher VOTs in males than in females, particularly around the age of ten. However, contrasting results on the effect of gender on VOT patterns were found in the literature (cf. Koenig, 2000, who reported relatively longer VOTs in females than males; Sweeting & Baken, 1982, who highlighted the lack of a gender effect on VOT measurements).

Against this background, which reveals discrepancies in the findings of previous studies on the production of VOTs of English stops and the effect of gender on the matter, the researcher asserts that the current study is required to offer more precise findings on Arabic-speaking ESL learners' production of English stops. Thus, the present paper investigated (i) if gender differences exist in the acquisition of VOT for the English bilabial stops /p/ and /b/ by measuring the VOT values produced by Arabic-speaking Saudi ESL learners; and (ii) if the VOT values for /p/ and /b/ of Saudi ESL learners differ from those of NESs. To the best of the researcher's knowledge, there is scant research on this topic in the Saudi Arabian ESL context. Thus, one of the focuses of this investigation was identifying if significant gender differences exist between VOT measurements for this cohort. The researcher then compared the VOT data with the ranges of NESs' VOT values to test the CA hypothesis. In particular, this research project sought to answer two main research questions:

- a) Are there statistically significant differences between the VOT values of Saudi Arabic-speaking males and females for the English bilabial stops /p/ and /b/?
- b) Do VOT values produced by Saudi Arabic-speaking males and females for /p/ and /b/ differ from those of native English speakers?

1.1. Research hypotheses

The study hypothesized the following:

- a) There are statistically significant differences between the VOT values of Saudi Arabic-speaking males and females for the English bilabial stops /p/ and /b/.
- b) The VOT values for /p/ and /b/ produced by Saudi males and females differ from those of native English speakers.

2. Methodology

2.1. Participants

A total of 20 Arabic-speaking ESL Saudi participants were recruited (10 females and 10 males). All participants were aged 20 to 29 to avoid the effect of age on the results (Xue et al., 2006; Yu et al., 2015). They share a relatively similar education level (undergraduates) as educational differences may affect VOT values (Flege et al., 1999). All participants have approximately the same exposure to English over a relatively similar period (6-7 years).

2.2. Stimuli

Under controlled conditions—namely a quiet room—the participants were asked to phonate /p/ and /b/ at a comfortable pitch and intensity. These phonemes exist in two meaningful words in English at the initial position of a CVC context. For the sound /p/, the participants were asked to pronounce the word *peen*. Moreover, to measure the phoneme /b/, they were asked to phonate the word *been*.

2.3. Procedures

Praat software was used to record the phonetic data at a sampling frequency of 44,100 Hz. The microphone-to-mouth distance was set at 10 cm. The recordings happened in a quiet venue. The participants were asked to produce each of the words (*peen* and *been*) a total of six times to provide a benchmark. The data were then analyzed using Praat speech analysis software by two raters, the experimenter and a research assistant, to ensure interrater reliability. For each participant, the VOT value was measured for each iteration of the six utterances. All measurements were agreed on by both raters. In the end, a total of 240 valid samples of the target phonemes were coded for further analysis.

The VOT is an acoustic measurement that determines “the duration of the [period] between the release of a [stop] and the onset of the vocal-fold vibration as measured in milliseconds (ms)” (Batuk & Kayikci, 2020, p. 1). It was introduced by Lisker and Abramson (1964) over a half-century ago (Cho et al., 2019). This acoustic measurement is relevant for investigating occlusive consonants’ characteristics, namely, to determine the voicing contrast between voiced and voiceless stops.

2.4. Data Analysis

The VOT values for the target voicing contrast (/p/ and /b/) were analyzed using descriptive statistics, namely, the means. To investigate gender differences in the participants’ VOT values, an independent samples *t*-test was used to examine the null hypothesis that there are no statistically significant differences between the VOT values of Saudi Arabic-speaking males and females for the English bilabial stops /p/ and /b/. This test compared the means of the VOT values for males and females. All statistical analyses in this investigation were conducted using IBM SPSS Statistics (Version 20) predictive analytics software.

To address the question of whether the VOTs for /p/ and /b/ produced by Saudi males and females differ from those of NESs, the VOTs of Saudis were compared to the ranges of VOTs produced by NESs—derived from secondary data—to test the CA hypothesis. Many researchers provide data on these VOT ranges (e.g., Auzou et al., 2000; Lisker & Abramson, 1964; Lin & Wang, 2011; Rae, 2018). Lisker and Abramson (1964) stated that the VOT value of /p/ for NESs ranges from 20 to 120 ms, with an average of 100 ms. Meanwhile, the VOT value of /b/ produced by NESs is typically less than 20 ms (Lin & Wang, 2011; Rae, 2018). Auzou et al. (2000) posited that it ranges from 0 to 25 ms.

For the purposes of the current investigation and in line with existing research (e.g., Auzou et al., 2000; Lisker & Abramson, 1964; Lin & Wang, 2011; Rae, 2018), the voicing contrast between male and female participants was operationally defined as the acoustic measurement of the participants’ VOT values. This acoustic dimension represents a measure that is used for such analyses.

2.5. Reliability

To assess consistency in VOT measurements, 15% of the data were randomly chosen and then reanalyzed by the researcher, using the same procedures, one month after the original VOT measurements were conducted. An analysis of A Pearson’s product-moment correlation revealed a significant result ($r = .98$; $p < .001$), suggesting a high level of measurement reliability.

Interrater reliability was evaluated based on the variation between the two sets of data independently measured by the two raters. Pearson’s correlation coefficient analysis revealed that the interrater reliability was $r = .91$, indicating a high interrater agreement.

3. Findings and Discussion

This research project aimed to examine gender-based differences in the acquisition of English bilabial stops among ESL Saudis. To achieve this, the researcher compared male and female participants' VOT values for /p/ and /b/. Moreover, this paper sought to address whether the VOT values for /p/ and /b/ produced by Saudis differ from those of NESs.

The next section presents an overview of the descriptive statistics for males' and females' VOT measurements. Next, it discusses the independent samples *t*-test results. Finally, it compares the VOT values for Arabic-speaking ESL Saudis versus the ranges defined for NESs mentioned above.

3.1. Descriptive statistics

Table 1 presents the means for the VOT values of /p/ and /b/ for male and female Saudi participants. It indicates that the male group had higher VOT values for /p/ ($M = 80.8$; $N = 10$) than females ($M = 78.7$; $N = 10$). Further, similar differences were found in the VOT values for /b/ where the scores for the males ($M = 27.2$; $N = 10$) were also higher than those of the females ($M = 16.1$; $N = 10$).

Table 1. Descriptive statistics for the VOT values (in Milliseconds)

	Gender	Statistic	Bootstrap ²	
			BCa 95% confidence interval	
			Lower	Upper
/p/ VOT value	Male	N	10	
		Mean	80.80	73.50 88.14
	Female	N	10	
		Mean	78.70	68.72 88.60
/b/ VOT value	Male	N	10	
		Mean	27.20	20.50 34.50
	Female	N	10	
		Mean	16.10	12.75 20.86

² Hereafter, bootstrap results in the current research are based on 10000 bootstrap samples.

3.2. T-test analysis results

An independent samples *t*-test was performed to examine gender-based differences in VOT values for /p/ and /b/. The results are summarized in Table 2 and discussed in the following subsections.

Table 2. Bootstrapped independent samples *t*-test of the mean VOT values

	Mean difference	Bootstrap			
		Bias	Std. Error	BCa 95% Confidence Interval	
				Upper	Lower
/p/ VOT value	2.100	.011	6.160	-9.331-	13.125
/b/ VOT value	11.100	-.021-	4.539	2.872	19.700

Using a percentile bootstrapping method, an independent samples *t*-test found no statistical difference between /p/ VOTs for the male group and those of the female group.

As Table 2 indicates, the 95% BCa CI for the difference in means was [-9.33, 13.12]. As the CI goes through zero, we conclude that there is no significant statistical difference between the examined groups.

For /b/ VOTs, the independent samples *t*-test found a statistical mean difference of 11.1 ms between the VOTs for the male group and those of the female group. The 95% BCa CI for the mean difference between the two groups was [2.87, 19.70], highlighting that there was at least a 3-point difference between males and females. This CI indicates that the mean difference between the examined groups, with 95% confidence, lies within this interval. Since the CI range does not contain zero, nor either its upper or lower bounds approach this value, this difference is considered statistically significant. Furthermore, a Cohen's *d* effect size for this comparison was 1.6 (using the SD of the female group [7] as a standardizer). According to Plonsky and Oswald's guidelines (2014), this represents a statistically large effect size.

Despite the *t*-test analysis results showing a statistical gender difference between the two means VOTs of /b/, this does not necessarily mean that the effect has real-life importance. Since the glottal cycle for consonants ranges from 7 to 8 ms, the statistical difference must be a minimum of two glottal cycles to be detected by the human ear. In this case, neither the lower bound (2.87 ms) nor the upper one (19.70) is sizable enough. Thus, the real-life effect size is negligible.

3.3. Discussion of the results

The findings of this research project showed that there is no significant difference between /p/ VOT values for male and female participants. This result concurs with Sweeting and Baken's (1982) finding that no gender differences exist for /p/ VOT values. Such similar results between males and females in their VOT values of /p/ are likely caused by the fact that the participants lack this phoneme in their native language. Additionally, all participants are late bilinguals who have had the same exposure to English—where /p/ is one of its phonemes—over a relatively similar period.

The findings also indicated that there is a significant difference between /b/ VOT values for male and female participants. This result concurs with Koenig's (2000) and Yu et al.'s (2015) findings where gender differences were evident in VOT values. In the current study, the male participants exhibited relatively higher VOT values than females, resembling Yu et al.'s (2015) result, while Koenig (2000) reported females having relatively higher VOTs.

Next, to investigate whether the VOTs for /p/ and /b/ produced by Saudi males and females differ from those of NESs, Saudis' VOTs were compared to the ranges of VOTs produced by NESs. As mentioned above, the VOT value of /p/ for NESs ranges from 20 to 120 ms (Lisker & Abramson, 1964). Therefore, the VOTs of /p/ for the male participants ($M = 80.8$) and the female participants ($M = 78.7$) were relatively similar to those of the English speakers. Further, the mean VOT of /b/ for NESs ranges from 0 to 25 ms (Auzou et al., 2000), while those of the Saudi male and female participants were 27.2 and 16.1 ms, respectively. Thus, while the mean for the female participants fell within the range for NESs, the males' mean was relatively higher. This higher mean VOT value for males is likely caused by the phonemes' position in the target words selected for the present study; in an isolated word, a phoneme presence in the initial position usually affects the VOT by increasing its value. With further testing that compares their (i.e., males') VOTs for /p/ and /b/, this higher VOT value could be investigated.

As mentioned, while Arabic features the phoneme /b/, it does lack the phoneme /p/. According to CA, such differences tend to cause difficulties for ESL learners. For example, Askoul (2017) found that learners tend to replace the phoneme /p/ with /b/. However, in the current study, despite that differences exist between English (as TL) and Arabic (as L1), they did not appear to cause difficulties

in acquiring the phoneme /p/. In fact, when the participants' VOTs were compared with the ranges, none of the participants phonated /p/ as /b/. These results concur with Abdelaal's (2017) work, wherein he reported that Arabic-speaking participants could phonate /p/ and /b/ in a relatively similar way to NESs. In the present study, the participants' acquisition of the voicing contrast is likely due to their advanced English level.

4. Conclusions

While some existing studies identified the existence of gender differences in VOT values (e.g., Koenig, 2000; Korkmaz & Boyaci, 2018), others suggested that no such gender differences in VOT values exist (e.g., Sweeting & Baken, 1982). The current results were aligned with the former for the phoneme /b/; namely, that, /b/ VOT values of male and female Saudis differed significantly. Meanwhile, a statistically significant difference was not found between /p/ VOTs of Saudi males and females. For the phoneme /b/, the results also showed a noticeably higher VOT value among the male participants compared to those of NESs. Other than this, the participants' overall VOT values in this study were relatively similar to those of NESs. These findings suggested that Saudi participants in this study have successfully acquired this voicing contrast, despite lacking the phoneme /p/ in their L1 (Arabic).

Despite its merits, the present research is subject to two main limitations. First, the presence of the target phonemes in word-initial position in the context of isolated words (not as part of a sentence where they may have been affected by catenation) might have increased the participants' VOT values. To overcome this, future research should investigate these phonemes in carrier sentences to provide more accurate results of gender differences in VOT values in real-life utterances. Second, the data on the VOTs of NESs used in the present study was based on secondary sources; therefore, future studies should sample NESs as a control group to provide more accurate data for comparison with the VOTs of L2 English speakers.

References

- Abdelaal, N. M. (2017). Instrumental analysis of the English stops produced by Arabic speakers of English. *International Journal of Education and Literacy Studies*, 5(3), 8-15. <https://doi.org/10.7575/aiac.ijels.v.5n.3p.8>
- Al-Shujairi, Y. B. J., Muhammed, A., & Almahammed, Y. S. O. (2015). Transitivity and intransitivity in English and Arabic: A comparative study. *International Journal of Linguistics*, 7(6), 38-52. <https://doi.org/10.5296/ijl.v7i6.8744>
- Askoul, S. (2017). An acoustic analysis of the production of word-initial stop /p/ by late Arab bilinguals. *TLC Journal*, 1(4), 74–90.
- Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics*, 14(2), 131–150.
- Batuk, I. T., & Kayikci, M. E. (2020). Evaluation of the voice onset time in Turkish-speaking schoolchildren. *International Journal of Pediatric Otorhinolaryngology*, 137, 1–5.
- Cho, T., Whalen, D., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, 72, 52–65. <https://doi.org/10.1016/j.jocn.2018.11.002>

- Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory & Language*, 41, 78-104.
- Koenig, L. L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research*, 43, 1211–1228.
- Korkmaz, Y., & Boyaci, A. (2018). Analysis of speaker's gender effects in voice onset time of Turkish stop consonants. *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, 1-5. <https://doi.org/10.1109/isdfs.2018.8355341>
- Lado, R. (1976). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.
- Lin, C., & Wang, H. (2011). Automatic estimation of voice onset time for word-initial stops by applying random forest to onset detection. *The Journal of the Acoustical Society of America*, 130(1), 514-525. <https://doi.org/10.1121/1.3592233>
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, 20(3), 384-422.
- Nofal, K. H. (2011). Nouns and adjectives of Old English and Modern Standard Arabic: A comparative study. *International Journal of Humanities and Social Science*, 1(18), 203–225.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912.
- Rae, R. (2018). *Measures of voice onset time: A methodological study*. [Master's thesis, College of Bowling Green State University]. Retrieved from <https://etd.ohiolink.edu/>
- Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Pearson Education Limited.
- Sweeting, P. M, Baken, R. J. (1982). Voice onset time in a normal-aged population. *Journal of Speech and Hearing Research*, 25, 129–134.
- Xue, S. A, Hao, G. J. P., & Mayo, R. (2006). Volumetric measurements of vocal tracts for male speakers from different races. *Clinical Linguistics & Phonetics*, 20(9), 691–702.
- Yu, V. Y., Nil, L. F. D., & Pang, E. W. (2015). Effects of age, sex and syllable number on voice onset time: Evidence from children's voiceless aspirated stops. *Language and Speech*, 58(2), 152–167. <https://doi.org/10.1177/0023830914522994>

AUTHOR BIODATA

Wasmiah Alsuheim is a Master's student of linguistics at the Department of English Language and Translation, College of Arabic Language and Social Studies in Qassim University, Saudi Arabia.