

The production of the English stop voicing contrast by Arab L2 speakers of English

Mohd Hilmi Hamzah^{1*}, Ahmed Elsayed Samir Madbouly¹, Hasliza Abdul Halim¹, and Abdul Halim Abdullah²

¹*School of Languages, Civilisation & Philosophy, UUM College of Arts & Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*

²*School of Education, Faculty of Social Sciences & Humanities, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*

ABSTRACT

The English voiceless stop /p/ and voiced stop /g/ are absent in the consonant inventory of Arabic. This difference provides a fertile ground for empirical research in L2 speech learning among Arab L2 speakers of English. The current study, therefore, aims to explore the English stop voicing contrast as produced by Arab native speakers. Focusing on Voice Onset Time (VOT) as an acoustic parameter, the study seeks to examine the extent to which (1) Arab L2 speakers of English maintain the English stop voicing contrast for /p-b/ and /k-g/, and (2) the L2 VOT continuum by Arab L2 speakers follows or deviates from the L1 VOT continuum in English. The acoustic phonetic experiment involved elicited materials of /p-b/ and /k-g/ from four male native speakers of Arabic. The tokens were recorded in isolation (utterance-initial position) and in a carrier sentence (utterance-medial position). The data were then acoustically analysed following standard segmentation, annotation and measurement criteria. Results reveal that the Arab L2 speakers can, to a large extent, maintain the English stop voicing contrast across all places of articulation, with voiced stops usually being produced with “normal” negative VOT (prevoicing) and voiceless stops usually being produced with “normal” positive VOT and also accompanied with aspiration in the long-lag region. There are also exceptional cases of “abnormal” negative VOT (prevoicing) for voiceless stops and “abnormal” positive VOT (devoicing) for voiced stops, with an extremely larger number of devoiced tokens for voiced stops in comparison to prevoiced tokens for voiceless stops. The results accord well with the Speech Learning Model’s prediction that phonetically “new” sounds are relatively easier to learn than phonetically “similar” sounds. The conclusion is drawn that languages sharing the same sound contrast may exhibit different phonetic implementations in marking a phonological contrast.

Keywords: Acoustic phonetics; Arab L2 speakers; L2 speech learning; stop voicing contrast; voice onset time

First Received:

16 October 2020

Revised:

21 February 2020

Accepted:

12 May 2020

Final Proof Received:

26 May 2020

Published:

30 September 2020

How to cite (in APA style):

Hamzah, M. H., Madbouly, A. E. S., Halim, H. A., & Abdullah, A. H. (2020). The production of the English stop voicing contrast by Arab L2 speakers of English. *Indonesian Journal of Applied Linguistics*, 10(2), 434-444. <https://doi.org/10.17509/ijal.v10i2.28615>

INTRODUCTION

A phonological contrast between voiced and voiceless stop consonants exists in many languages, such as the English voiceless stops /p, t, k/ versus

their voiced counterparts /b, d, g/. This particular contrast has been a major topic of investigation in phonetics and phonology over the last few decades. In the fields of second language (L2) acquisition and

* Corresponding Author
Email: hilmihamzah@uum.edu.my

L2 speech learning, many researchers have conducted various studies focusing on the pronunciation difficulties of English as an L2 (e.g., Calvo, 2016; Carlet & Rato, 2015; Derwing & Munro, 2015; Everitt, 2015; Ghulamullah & Hamzah, 2017; Hamzah, Ahmad, & Yusuf, 2017; Hamzah & Bawodood, 2019; Lee & Hwang, 2016; Lerdpaisalwong, 2015; Naser & Hamzah, 2018).

As for Arab L2 speakers of English, the focus of the current study, they usually encounter difficulties when pronouncing /p/ and /g/ in English as these two phonemes do not exist in the Arabic consonant inventory (see also Abdelaal, 2017; Al-Ahdal et al., 2015; Al-Malwi, 2017; AlTamimi, 2015; Huwari & Mehawesh, 2015; Ismael, Mahadin, & Masri, 2015; Sabbah, 2015; Shormani & Alsohbani, 2015). Previous studies have shown that, among Arab L2 speakers, the voiceless bilabial stop /p/ was often substituted by its voiced counterpart /b/ that is available in the Arabic phonemic inventory, such as “park” potentially produced as “bark” (Buali, 2010; Flege & Port, 1981; Khattab, 2002). However, the voiced velar stop /g/, and the potential replacement with its voiceless counterpart /k/, is not getting much attention from researchers in this area. The current study fills this gap by focusing on the production of /p/ and /g/ by Arab L2 speakers of English.

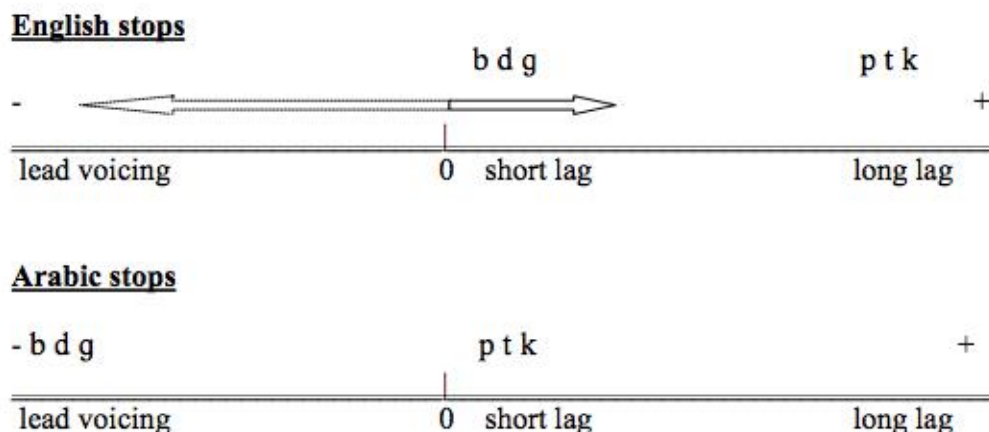
Many theories and models concerning sounds have been proposed in order to understand how and why L2 speakers, particularly adult speakers, usually differ from native speakers. Speech Learning Model (SLM) of Flege (1980), the chosen model of the current study, takes a holistic view of differences between languages and their effect on L2 acquisition. In this model, Flege (1980) places a

great emphasis on the Voice Onset Time (VOT), which is one of the important differences between the English and Arabic stops. VOT is a key feature of stop voicing contrasts that reflects the period between the stop release and the beginning of a vowel (Lisker & Abramson, 1964). In languages with stop voicing contrasts, such as in English and Arabic, VOT is the critical acoustic phonetic property that marks the differences between the two languages (see also Hamzah, Fletcher, & Hajek, 2011). These differences have provided a basis for empirical research in L2 speech learning among Arab L2 speakers of English.

Stop voicing contrast: English vs. Arabic

English and Arabic stop voicing contrasts can be differentiated on the basis of their phonetic implementation (Khattab, 2002). As indicated in Figure 1, aspirated voiceless stops in English are produced in the long-lag region (long positive VOT) and voiced stops are produced in the short-lag region (zero or short positive VOT) (Deuchar & Clark, 1996; Lisker & Abramson, 1967; Mahmood, 2016). Unaspirated voiceless stops in English are also produced in the short-lag region. Prevoicing (negative VOT) is uncommon in English stops. Note that, in English, voiced stops are often not phonetically voiced (i.e., they are not accompanied by vocal cord vibration). Generally, most native speakers of English do not voice initial voiced stops to the full extent. The stop voicing contrast in Arabic, on the other hand, is characterised by short lag (zero or short positive VOT) for voiceless stops, and prevoicing (negative VOT) for voiced stops (Flege & Port, 1981).

Figure 1
VOT Continuum between English and Arabic Stops



Reprinted from “VOT production in English and Arabic bilingual and monolingual children”, by G. Khattab, 2000, *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 8, p. 96. Copyright 2000 by Ghada Khattab.

Based on the comparison above, Arabic can be considered to be a voicing language and English an aspiration language, as supported by Kager et al.

(2007). That is to say, although both languages show a distinction between stops labeled phonologically as voiced versus voiceless, Arabic

stops differ phonetically via voicing while English stops are distinguished by aspiration (Ladefoged & Maddieson, 1998). Despite the VOT differences, stop voicing contrasts in distinct languages such as English and Arabic always share a similar phonological function and differ in the phonetic implementations employed to show such contrasts.

There have been some L2 studies that examined the production of VOT by Arab L2 speakers of English (e.g., Buali, 2010; Flege, 1980; Flege & Port, 1981; Khattab, 2002; Port & Mitleb, 1983). Overall, these studies reported similar results: although both English and Arabic languages share similar stop categories, they exhibit different VOT patterns for such categories.

Speech Learning Model

The SLM hypothesises that L2 learners perceive certain L2 sounds as “new”, while some other sounds are thought to be “similar” to the learners’ L1. For example, for Arab L2 speakers, the English /p/ is “new” since there is no such a sound in Arabic, while /b/ is “similar” since both Arabic and English have this particular sound in their phonemic inventory. Note that, there are considerable VOT differences for the “similar” sound /b/ in Arabic and English. As discussed earlier, the Arabic /b/ is normally prevoiced, while the English /b/ is usually produced in the short-lag region (positive VOT).

According to Flege (1995), “new” or “similar” sounds are based on the allophonic properties of such sounds, not on their phonemic properties. Flege (1980) also asserts that, at the initial stage of L2 learning, L2 learners are more receptive to “similar” L2 sounds in comparison to “new” L2 sounds. However, as L2 learning progresses, the pattern is reversed: the learners become more attuned to “new” than “similar” L2 sounds, thus achieving a higher level of accuracy in the production and perception of “new” sounds rather than on “similar” sounds.

Purpose of the study

The purpose of the current study is to investigate how well Arab L2 speakers acquire the VOT of English stops and to what extent these speakers transfer phonetic values for the newly acquired language from their native language. Experimental findings of the current study will test the prediction of the SLM concerning these situations. The findings will have important implications in the areas of L2 speech learning as well as pronunciation teaching and learning. More importantly, the findings of the current study will fill the knowledge gap in the phonetic literature concerning VOT research and speech science.

This study aims to answer the following research questions:

1. To what extent do Arab L2 speakers maintain a voicing contrast for /p-b/ and /k-g/ in their

L2 English in terms of VOT? Does /p/ show greater “abnormal” negative VOT (prevoicing)? Does /g/ show greater “abnormal” positive VOT (devoicing)?

2. To what extent does the English stop voicing contrast produced by Arab L2 speakers follow or deviate from the VOT continuum in English?

Two hypotheses are put forward in this study. First (Hypothesis H₁), it is hypothesised that Arab L2 speakers do not maintain a voicing contrast for /p-b/ and /k-g/ in their L2 English in terms of VOT, with a greater number of “abnormal” negative VOT (prevoicing) for /p/ and a greater number of “abnormal” positive VOT (devoicing) for /g/. Second (Hypothesis H₂), it is hypothesised that the English stop voicing contrast produced by Arab L2 speakers deviates a great deal from the VOT continuum in English. Following SLM, it is also expected that “new” L2 sounds (/p/ and /g/) are much easier to produce than “similar” L2 sounds (/b/ and /k/).

METHOD

Materials

An acoustic phonetic experiment was designed to investigate the English stop voicing contrast produced by Arab L2 speakers of English. A list of sixteen tokens was prepared consisting of eight minimal pairs (presented in Table 1). All tokens were chosen so as to provide the word-initial stop voicing contrast in English. They were all English monosyllabic words with the CVC structure. Two types of place of articulation were chosen and they were grouped according to voicing profile: (1) the voiceless bilabial stop /p/ versus the voiced bilabial stop /b/; and (2) the voiceless velar stop /k/ and the voiced velar stop /g/. In order to control the vowel effect, each stop was followed by either the high front vowel /i/ or the low central vowel /a/. All tokens were familiar words and well known to the participants. All target sounds were located in word-initial position; it is well known that onsets provide a better environment to examine speech production (see, e.g., Hamzah, Hajek, & Fletcher, 2020).

Speakers

The participants were four native speakers of Arabic whose ages ranged between 24 to 29 (mean age: 26.5). For cultural reasons, only male participants were recruited. They were all students from a university located in the state of Kedah, Malaysia. Three of them were undergraduate students, while one of them was a PhD student. The speakers were selected through the second author’s personal contacts. At the time of the experiment, they had been studying at the university for between one to two years. They rated themselves as good English

language speakers. Two of the speakers were born in Yemen, one speaker in Chad and the other one in Somalia. Although they came from different nationalities, it was confirmed by the second author

(who is also a native speaker of Arabic) that all participants were native Arabic speakers with a similar level of nativeness.

Table 1

List of Tokens and Their International Phonetic Alphabet (IPA) Transcriptions

Place of articulation	Vowel height	Voiceless stops	IPA	Voiced stops	IPA
Bilabial	/i/	pin	/pɪn/	bin	/bɪn/
		pill	/pɪl/	bill	/bɪl/
	/a/	pun	/pʌn/	bun	/bʌn/
		puff	/pʌf/	buff	/bʌf/
Velar	/i/	kill	/kɪl/	gill	/gɪl/
		kid	/kɪd/	gid	/gɪd/
	/a/	cut	/kʌt/	gut	/gʌt/
		come	/kʌm/	gum	/gʌm/

Data collection

The experimental materials were recorded individually in a soundproof laboratory using a professional Sony recorder. In all sessions, speakers were asked to produce each token in two different utterance contexts: (1) in isolation (i.e., utterance-initial position); and (2) in a carrier sentence (i.e., utterance-medial position). The first context required a long silent pause after the target word, while the second context required a vowel after the target word. The carrier sentence used in the study was “I say (the target word)”, which was adopted from Khattab’s (2002) study. The carrier sentence was written separately on a piece of A4 paper.

All experimental tokens were presented in random order using a PowerPoint slide presentation on a computer. Ten distractors were used to reduce the participants’ awareness of the study’s research questions. The distractors were as follows: “like”, “rate”, “say”, “zoo”, “ship”, “hood”, “fair”, “name”, “make”, “jug”. These distractors were excluded from data analysis. The speakers used a headphone throughout the recording session. The tokens (and also the distractors) were shown three times to the speakers using three different lists, which were also randomised. Speech rate was not controlled, so the tokens were produced according to the speakers’ normal speech rate. The speakers went through an initial training in which they produced a number of tokens so that they were familiar with the procedures. The second author explained all the procedures to the speakers in the speakers’ native language. At the end of the experiment, each speaker produced 96 utterances in both utterance contexts, yielding 384 utterances for the whole L2 corpus of the English stop voicing contrast. The experiment took approximately one hour for each speaker.

Data analysis

The audio files were digitised at 44.1 kHz. They were later segmented into single utterances and coded accordingly for each speaker. Praat version 6.0.28 (Boersma & Weenink, 2017) was used for

segmenting and annotating in which the boundaries of segments were manually based on visual inspection of spectrographic and waveform information displayed in Praat. The procedures of segmenting and labeling voiceless stops and voiced stops were based on established criteria used in many acoustic phonetic studies (e.g., Croot & Taylor, 1995).

Figure 2 displays a set of waveforms and spectrograms illustrating the annotation of voiceless stop tokens produced in a carrier sentence. For the purpose of this study, only words and VOTs were labelled (this also applied to tokens beginning with voiced stops). As observed in Figure 2, three annotation tiers were derived in the Praat TextGrid: (1) the word tier (top tier); (2) the VOT tier (second tier); and (3) the remarks tier (third tier). The word tier (top tier) shows the segmentation and labeling of the target word (e.g., ‘kill’ as shown in Figure 2). The VOT tier (second tier) highlights either positive VOT (marked as ‘+h’) or negative VOT (marked as ‘-h’). The remarks tier (third tier) labels (1) the voiceless stop segments that were partially voiced (‘PV’) or fully voiced (‘FV’), or (2) the voiced stop segments that were partially devoiced (‘PD’) or fully devoiced (‘FD’). All the VOTs for partially/fully voiced segments of voiceless stops were considered as “abnormal” negative VOT (prevoicing) and labelled as ‘-h’ in the VOT tier. Contrariwise, all the VOTs for partially/fully devoiced segments of voiced stops were considered as “abnormal” positive VOT (devoicing) and labelled as ‘+h’ in the VOT tier.

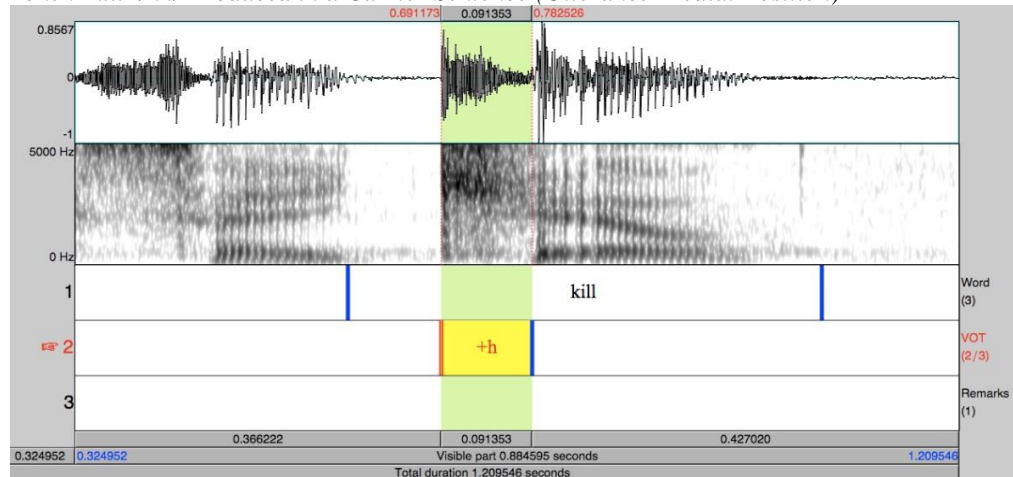
Following Lisker and Abramson (1964), positive VOT for voiceless stop tokens was measured in ms over the periods of the release of stops in both utterance contexts (marked as ‘+h’ in Figure 2). The measurement of negative VOT for voiced stop tokens usually corresponded to the closure duration measurement of voiced stops (see, e.g., Hamzah, 2010; Hamzah, Fletcher, & Hajek, 2016). That is, negative VOT was calculated utterance-initially from the onset of prevoicing to the release of the stop. Utterance-medially, it was

measured as the period of voicing throughout the closure phase until the onset of the release phase (marked as ‘-h’). If the burst was missing, the

endpoint was defined as the onset of voicing of the following vowel.

Figure 2

Annotated Waveform and Spectrogram in the Praat Textgrid Spoken by the Male Speaker, Speaker 1, from the Token ‘kill’ /kɪl/ Produced in a Carrier Sentence (Utterance-Medial Position)



With regard to the short/long-lag region of positive VOT, the estimate of Cho and Ladefoged (1999), and Khattab (2002) was adopted in this study: (1) short-lag region=0 to 30 ms; and (2) long-lag region=above 30 ms. As for aspiration, the following VOT categories were used (based on Cho & Ladefoged, 1999):

1. Unaspirated=0 to 30 ms
2. Slightly aspirated=30 to 50 ms
3. Aspirated=50 to 90 ms
4. Highly aspirated=above 90 ms

To describe the contrast maintenance found in this study, the “normal” and “abnormal” tokens were calculated based on the following criteria:

1. “Normal” tokens=positive VOT of voiceless stops, and negative VOT of voiced stops
2. “Abnormal” tokens=negative VOT of voiceless stops (prevoiced tokens), and positive VOT of voiced stops (devoiced tokens)

To describe the VOT continuum found in this study, the following symbols are used:

1. +/p/ and +/k/=voiceless stops produced with “normal” positive VOT
2. -/b/ and -/g/=voiced stops produced with “normal” negative VOT
3. -/p/ and -/k/=prevoiced tokens produced with “abnormal” negative VOT
4. +/b/ and +/g/=devoiced tokens produced with “abnormal” positive VOT

The statistical analyses, such as samples paired *t*-tests, were conducted on the VOT data to test the level of significant VOT differences between voiceless stops and voiced stops. The *p*-values at 0.05 or below were considered significant.

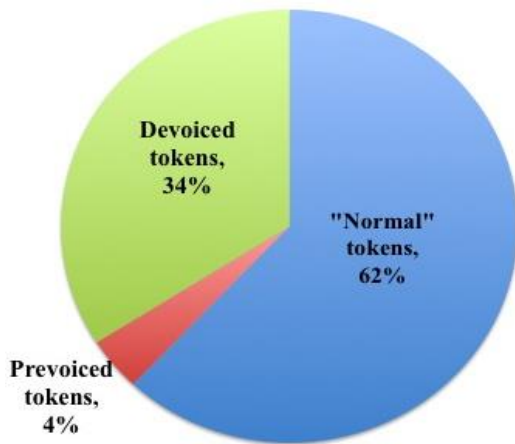
FINDINGS

Contrast maintenance

The number of “normal” and “abnormal” tokens is illustrated in Figure 3 (across the whole corpus) and Figure 4 (according to each speaker). Table 2 provides details that underlie these two figures. It can be seen that 62% of the tokens (i.e., 240 out of 384 tokens) are produced with “normal” positive/negative VOTs, with Speaker 1 producing the highest number of “normal” tokens (77 tokens), followed by Speaker 3 (60 tokens), Speaker 2 (55 tokens), and Speaker 4 (48 tokens). Only 4% of the voiceless stop tokens (i.e., 14 out of 384 tokens) are produced with “abnormal” negative VOT (prevoicing), with a greater number of prevoicing cases in tokens beginning with /p/ (10 tokens) than those beginning with /k/ (4 tokens only). In this case, Speaker 1 tends to prevoice the most (8 tokens), and followed by Speaker 3 (6 tokens). Speaker 2 and Speaker 4 do not attempt to prevoice at all.

With regard to devoicing, 34% of the voiced stop tokens (i.e., 130 out of 384 tokens) are produced with “abnormal” positive VOT, with /b/ having a larger number of devoicing cases (70 tokens) as compared to /g/ (60 tokens). In this context, Speaker 4 contributes the largest number of devoiced tokens (48 tokens), followed by Speaker 2 (41 tokens), Speaker 3 (30 tokens) and Speaker 1 (11 tokens). Most of “abnormal” negative VOTs (prevoicing) and “abnormal” positive VOTs (devoicing) are found in utterance-medial position (prevoiced tokens=10, devoiced tokens=68) than in utterance-initial position (prevoiced tokens=4, devoiced tokens=62) (see Table 2). VOT values are also usually greater in the /i/ environment than in the /a/ environment.

Figure 3
The Number of "Normal" Tokens and "Abnormal" Tokens (Prevoiced and Devoiced Tokens) Across the Whole Corpus (in Percentages)



VOT continuum

Figure 5 demonstrates the VOT continuum for "normal" VOTs against "abnormal" VOTs produced by each speaker in this study. Table 3 provides the detailed measurements. Each mean VOT value reported in Table 3 was measured across vowel and utterance contexts. In general, it can be observed in Figure 5 that most of the VOTs are located in the left continuum (negative VOT) and also in the right continuum (positive VOT), except for Speaker 4 in which only positive VOTs are produced (both "normal" and "abnormal" positive VOTs). For other speakers (Speaker 1, Speaker 2, and Speaker 3), almost all negative VOTs are fully voiced, i.e., a negative value above -75 ms, as outlined by Lisker and Abramson (1964). The longest stop that is fully

voiced is the "abnormal" -/p/ (-114 ms) produced by Speaker 1.

As for positive VOTs, they are mostly aspirated and produced within the long-lag region, i.e., above 30 ms, based on Cho and Ladefoged (1999), and Khattab (2002). The most aspirated stop is the "normal" +/p/ (87 ms) produced by Speaker 4. Note that this VOT value (i.e., 87 ms) is close to the "highly aspirated" category (i.e., above 90 ms). For most speakers, the "abnormal" negative VOT values for -/b/ and -/g/ are greater than the "normal" ones for -/b/ and -/g/. On the contrary, the "normal" positive VOT values for +/p/ and +/k/ are always greater than the "abnormal" ones for +/b/ and +/g/.

The VOT continuum for Speaker 4 requires some additional notes. The most striking pattern for this particular speaker is that there is no negative VOT involved in the production of the English stop voicing contrast. That is, all stop tokens are produced with positive VOTs, including the voiced stop tokens (which are all fully devoiced). There is no prevoiced case for voiceless stops. It is also worth remarking that the VOT for the "normal" +/p/ is longer than that for the "normal" +/k/, which is in contradiction with the expected universal trend for place of articulation (i.e., velars have longer VOTs than bilabials). Note also for Speaker 4 that there is a stark contrast between "abnormal" and "normal" positive VOTs. On one hand, the "abnormal" positive VOTs are all unaspirated (in the short-lag region), while on the other hand, the "normal" positive VOTs are all aspirated. The mean VOT differences between stop contrasts are all large (i.e., 57 ms for +/p/ vs. +/b/, and 49 ms for +/k/ vs. +/g/) and highly significant (all $p < 0.001$).

Figure 4
The Number of "Normal" Tokens and "Abnormal" Tokens (Prevoiced and Devoiced Tokens) According to Speaker (the Data were Pooled Across Vowel and Utterance Contexts)

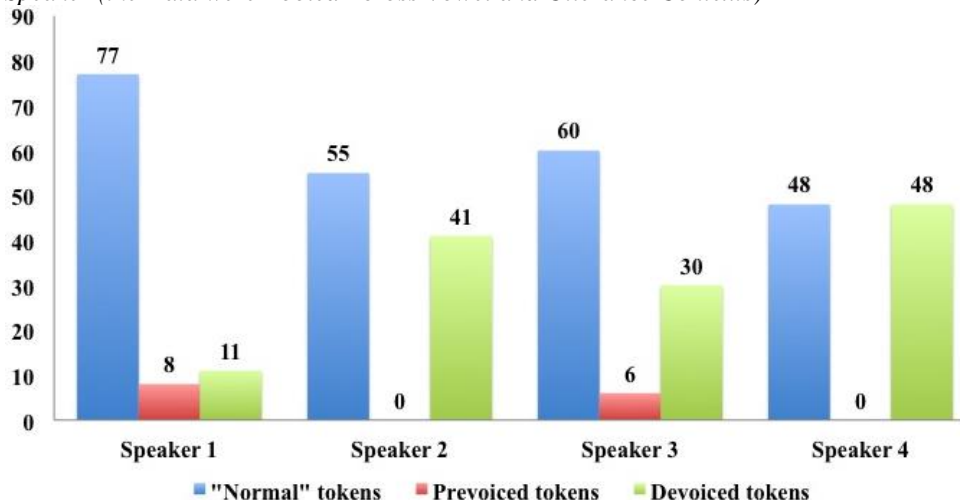


Table 2

The Number of “Normal” Tokens and “Abnormal” Tokens (Prevoiced and Devoiced Tokens) According to Speaker and Utterance Context (‘Ui’ Utterance-Initial Position; ‘Um’ Utterance-Medial Position)

Tokens	Speaker 1	Speaker 2	Speaker 3	Speaker 4	TOTAL
“Normal”	Ui=42 Um=35	Ui=29 Um=26	Ui=31 Um=29	Ui=24 Um=24	Ui=126 Um=114
Total	77	55	60	48	240 (62%)
Prevoiced	Ui=1 Um=7	Ui=0 Um=0	Ui=3 Um=3	Ui=0 Um=0	Ui=4 Um=10
Total	8 /p/=4 /k/=4	0	6 /p/=6	0	14 (4%) /p/=10 /k/=4
Devoiced	Ui=5 Um=6	Ui=19 Um=22	Ui=14 Um=16	Ui=24 Um=24	Ui=62 Um=68
Total	11 /b/=4 /g/=7	41 /b/=20 /g/=21	30 /b/=22 /g/=8	48 /b/=24 /g/=24	130 (34%) /b/=70 /g/=60

Figure 5

VOT Continuum (ms) for “Normal” VOTs (/+p/, +/k/, -/b/, -/g/) and “Abnormal” VOTs (-/p/, -/k/, +/b/, +/g/) According to Speaker. “Abnormal” VOTs are Highlighted in Red.

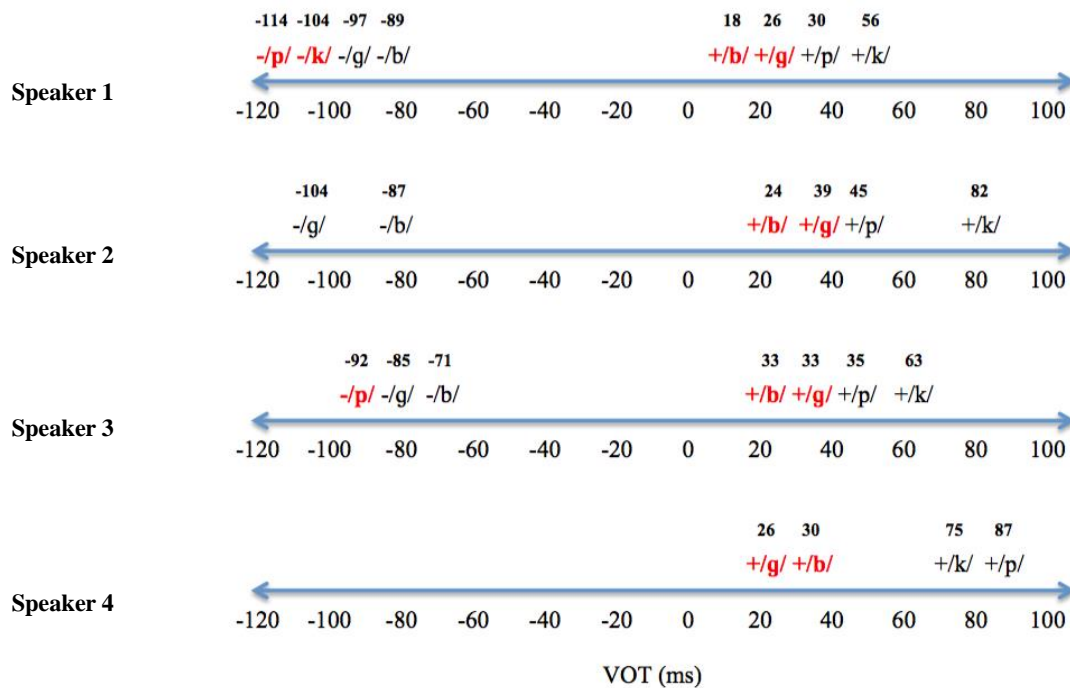


Table 3

Details of VOT Continuum (ms) for “Normal” VOTs (/+p/, +/k/, -/b/, -/g/) and “Abnormal” VOTs (-/p/, -/k/, +/b/, +/g/) According to Speaker. “Abnormal” VOTs are Highlighted in Red.

	Speaker 1			Speaker 2			Speaker 3			Speaker 4		
	N	Mean	Sig.	N	Mean	Sig.	N	Mean	Sig.	N	Mean	Sig.
+p/	20	30	<0.01	24	45	<0.001	18	35	<0.01	24	87	<0.001
+k/	20	56		24	82		24	63		24	75	
-b/	20	-89	<0.01	4	-87	<0.05	2	-71	<0.05	0	-	-
-g/	17	-97		3	-104		16	-85		0	-	
-p/	4	-114	<0.05	0	-	-	6	-92	-	0	-	-
-k/	4	-104		0	-		0	-		0	-	
+b/	4	18	<0.05	20	24	<0.01	22	33	n.s.	24	30	<0.001
+g/	7	26		21	39		8	33		24	26	

DISCUSSION

Maintenance of the English stop voicing contrast by Arab L2 speakers

The findings reported in this study partially support Hypothesis H₁: all Arab L2 speakers can, to a great extent, maintain the English stop voicing contrast using VOT (62% of the tokens). In “normal” tokens, most English voiceless stops are produced with aspiration in the long-lag region, while English voiced stops are produced with prevoicing. With regard to “abnormal” tokens, there is an imbalance between prevoicing for voiceless stops and devoicing for voiced stops: the latter occurs more frequently (i.e., 34%) than the former (i.e., 4%). These results support Lisker and Abramson’s (1964) prediction that L2 speakers of English are able to distinguish voiceless stops from their voiced counterparts, albeit different VOT ranges from English, as also found among Thai and Dutch speakers (Lisker & Abramson, 1964; Simon, 2009) who both lack the voiced velar stop /g/ in their L1s (like the speakers in the current study). It might be the case that the Arab L2 speakers in the present study transfer their L1 prevoicing to their English L2 voiced stops.

The findings in the current study are also in accordance with earlier studies examining L2 English among Arab speakers (Alves & Zimmer, 2015; Buali, 2010; Calvo, 2016; Flege & Port, 1981; Gordon & Darcy, 2016; Khattab, 2002; Port & Mitleb, 1983; Olson, 2017; Olson & Hayes-Harb, 2019; Rato & Rauber, 2015). In these studies, it was reported that all Arab L2 speakers are able to distinguish between the English voiced /b/ and the aspirated voiceless /p/, although this contrast does not exist in their L1 (Arabic). The “normal” positive VOT ranges reported for voiceless stops in these studies are all within the short-lag region, which runs counter to the VOT range reported in the current study for voiceless stops (i.e., the long-lag region). Some of these studies (Buali, 2010; Flege & Port, 1981) report that many /p/ productions have some “abnormal” negative VOTs (prevoicing) during the period of the stop closure, which is similar to the voiceless stop VOT data in the current study (4% of prevoicing).

As for voiced stops, some studies (e.g., Khattab, 2002) also report “normal” negative VOT (prevoicing) for voiced stop tokens, which reflects L1 transfer (i.e., voiced stops in Arabic are usually prevoiced). However, the “abnormal” positive VOT (devoicing) for voiced stops are not reported in any of these studies, unlike the current study in which 34% of voiced stop tokens are devoiced and produced with “abnormal” positive VOTs. Most of the devoiced tokens for voiced stops in the current study are successfully distinguished from their voiceless counterparts. That is, the “abnormal” positive VOT values for devoiced tokens are lower than the “normal” positive VOT values for voiceless

stops. This situation reflects a unique VOT strategy among Arab L2 speakers in this study in distinguishing voiced and voiceless stops, mirroring the VOT pattern among native speakers of English.

It appears that, although most Arab L2 learners of English reported in earlier and current studies successfully produce the English stop voicing contrast, they often substitute small phonetic details of an L2 with those of their L1. In this case, they produce the Arabic VOT pattern in their English production (e.g., “normal” negative VOT for voiced stops), lending evidence that both English and Arabic languages share some similar stop categories, but these two languages contrast in their VOT patterns.

English VOT vs. Arabic VOT

The results found in this study partially support Hypothesis H₂. First, the results on “normal” negative VOT (prevoicing) support Hypothesis H₂: most of the voiced stop tokens are produced with “normal” negative VOT (fully voiced most of the time), as shown in the VOT continuum for each speaker in this study, although many voiced stop tokens (34%) are also produced with “abnormal” positive VOT, particularly Speaker 4. The finding on “normal” negative VOT (prevoicing) generally contradicts with the VOT pattern in English, in which English voiced stops are not usually prevoiced. Instead, they are produced with shorter positive VOTs within the short-lag region. The VOT pattern shown by most Arab L2 speakers in the current study reflects the VOT pattern in their L1, in which voiced stops are usually produced with prevoicing. This has been shown to be case in many earlier VOT studies among Arab speakers in their L1 productions (e.g., Adam, 2012; Alghamdi, 1990; Flege, 1980; Yeni-Komshian, Caramazza, & Preston, 1977).

Second, on the contrary, the positive VOT results found in the current study do not support Hypothesis H₂: most Arab L2 speakers produce positive VOTs for voiceless stops with aspiration within the long-lag region, which is similar to the VOT pattern in English. Furthermore, the “abnormal” positive VOTs are produced within the short-lag region, which again mirrors the VOT pattern used by native speakers of English for voiced stops (and also for the unaspirated voiceless stops). That is, in English, phonologically termed voiced stops are often not phonetically voiced. A similar VOT pattern among Arab L2 speakers has also been observed in, for example, Aldahri (2013).

It seems that, in the first case (“normal” negative VOT in voiced stops), the Arab L2 speakers in this study may “carry over” the Arabic phonetic features of the stop voicing contrast onto their English voiced stop production. As for the second case (aspiration in voiceless stops), it seems that the Arab L2 speakers in the current study

manage to control all the articulatory dimensions (i.e., the glottal-supraglottal timing) typically used for the production of English aspirated voiceless stops. That is, they manage to produce native sounds that might be harder for most non-native speakers of English to produce, as shown in many L2 studies (e.g., Flege & Port, 1981). This can be explained using the theory of SLM.

Interpretations from the Speech Learning Model

This study has shown that most Arab L2 speakers do not distinguish between L2 voiced stops (with short lag) and their L1 voiced stops (with prevoicing). Based on Flege's (1987) SLM, it can be claimed that the English short-lag VOT is *similar* to Arabic prevoicing, which leads to the creation of an equivalence classification (or *assimilation*) for both types of sounds and consequently hinders Arab L2 speakers from forming a "new" phonetic category for L2 voiced stops that are produced with short-lag VOTs. In the case of Arab L2 speakers in this study, this situation causes them to "maintain" their L1 prevoicing pattern in their L2 speech, i.e., producing voiced stops with "normal" negative VOT (prevoicing).

With respect to voiceless stops, the Arab L2 learners in this study manage to perceive the difference between L2 long-lag voiceless stops (typically produced with aspiration) and their L1 short-lag stops (without aspiration). In the view of SLM, the English stop aspiration (which is not available in Arabic) is highly *dissimilar* from L1 unaspirated stops in Arabic, which enables the Arab L2 learners to create a "new" category for the L2 sound, i.e., English aspiration. That is, Arab L2 learners can produce aspiration because of *dissimilation* between sounds (more specifically, a salient acoustic difference between the aspirated long-lag VOT and the unaspirated short-lag VOT).

CONCLUSION

The primary goal of this study was to explore the production of the English stop voicing contrast by Arab L2 speakers by using VOT as an acoustic parameter. The findings reveal that all Arab L2 speakers of English recruited in this study can extensively maintain the English stop voicing contrast for /p-b/ and /k-g/ in terms of VOT, although the number of devoicing cases for voiced stops is extremely larger than that of prevoicing cases for voiceless stops. Note that, in this study, "abnormal" positive VOT (devoicing) is usually produced in the short-lag region, while "abnormal" negative VOT is long and fully prevoiced. That "abnormal" positive VOT (devoicing) frequently occurs for the voiced stop /g/ is expected, given the absence of this particular phoneme in the Arabic consonant inventory. As for the unexpected case of devoicing for the voiced stop /b/, which is available

in Arabic, this could be due to the speakers' attempt to mirror the native production of /b/, which is produced in English with positive VOTs in the short-lag region.

The study has some theoretical implications, particularly with regard to the SLM. Most of the voiceless stop tokens in this study are produced with aspiration, while many voiced stop tokens are devoiced and produced with short positive VOTs. One can still argue whether this linguistic phenomenon can be associated with an equivalence classification (assimilation) or a new category formation (dissimilation). It might be the case that there is another category in between, which is possible given the continuous state of L2 learning. Therefore, L2 speech learning model should also consider this possibility in its description of "new" and "similar" sounds.

Based on the results reported in this study, it can be concluded that Arab L2 speakers can, to a large extent, produce the English phonological contrast of stops, albeit different phonetic implementations of VOT in comparison to the native L1 norm. As such, this study has experimentally tested the role of VOT in characterising L2 stop contrasts among Arab L2 speakers and explored the potential cases of prevoicing and devoicing for voiceless stops and voiced stops, respectively. More broadly, it has contributed to the phonetic literature concerning VOT and, more specifically, L2 speech learning. In doing so, it appears that there are many other avenues for further studies in the L2 acquisition of the stop voicing contrast in English and also Arabic. For example, future researchers may further examine the articulatory differences between short-lag voiced stops and prevoiced stops that always cause some production difficulties among Arab L2 speakers of English. It is hoped that this study will be seen as a significant contribution to the fields of acoustic phonetics and applied linguistics.

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>
- Adam, H. (2012). VOT-analysis: The production of stops by agrammatical Palestinians. *International Journal of Linguistics, 4*(4), 300–308. <https://doi.org/10.5296/ijl.v4i4.2736>
- Al-Ahdal, A. A. M. H., Al-Hattami, A. A., Al-Awaid, S. A. A., & Al-Mashaqba, N. J. H. (2015). Pronunciation for the Arab learners of EFL: Planning for better outcomes. *English Language Teaching, 8*(10), 100–106. <https://doi.org/10.5539/elt.v8n10p100>

- Aldahri, S. S. (2013). A study for the effect of the emphaticness and language and dialect for Voice Onset Time (VOT) in Modern Standard Arabic (MSA). *Signal & Image Processing: An International Journal*, 4(2), 11–29. <https://doi.org/10.5121/sipij.2013.4202>
- Alghamdi, M. (1990). *Analysis, synthesis and perception of voicing in Arabic*. University of Reading.
- Al-Malwi, I. M. (2017). *The effects of gender and age on voice onset time by Abah Arabic speakers* [Unpublished bachelor doctoral dissertation]. Proquest Database. <https://pqdtopen.proquest.com/doc/1870036550.html?FMT=ABS&pubnum=10252690>
- Altamimi, A. K. (2015). *Improving English pronunciation among Arabic EFL school-age student using minimal pairs* [Unpublished master's thesis]. State University of New York. <https://dspace.sunyconnect.suny.edu/handle>
- Alves, U., & Zimmer, M. (2015). Perception and production of English VOT patterns by Brazilian learners: The role of multiple acoustic cues in a DST perspective. *Alfa*, 59(1), 155–175. <https://doi.org/10.1590/1981-5794-1502-7>
- Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Computer program], Version 6.0.28, Retrieved from <http://www.praat.org/>
- Buali, I. (2010). *The perception and production of /p/ in Saudi Gulf Arabic English: A variationist perspective* [Unpublished doctoral dissertation]. Concordia University. <https://www.semantic.scholar.org/paper/The-Perception>
- Calvo, Y. J. (2016). *The teaching and learning of English pronunciation in Spain: An analysis and appraisal of students' and teachers' views and teaching materials* [Unpublished doctoral dissertation]. University of Santiago de Compostela. <https://dialnet.unirioja.es/servlet>
- Carlet, A., & Rato, A. (2015). Non-native perception of English voiceless stops. In E. Babatsouli, & D. Ingram (Eds.), *Proceedings of the international symposium on monolingual and bilingual speech 2015* (pp. 57–67). Institute of Monolingual and Bilingual Speech.
- Cho, T., & Ladefoged, P. (1999). Variation and universal in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27(1), 207–227. <https://doi.org/10.1006/jpho.1999.0094>
- Croot, K., & Taylor, B. (1995). *Criteria for acoustic-phonetic segmentation and word labelling in the Australian national database for spoken language*. Macquaire University. http://andosl.anu.edu.au/andosl/general_info/ae_criteria.html
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company.
- Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, 24(3), 351–365. <https://doi.org/10.1006/jpho.1996.0019>
- Everitt, C. (2015). *Accent imitation on the L1 as a task to improve L2 pronunciation* [Unpublished master's thesis]. University of Barcelona. <http://hdl.handle.net/2445/69627>
- Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, 30(1), 117–134. <https://doi.org/10.1111/j.1467-1770.1980.tb00154.x>
- Flege, J. E. (1987). The instrumental study of L2 speech production: Some methodological considerations. *Language Learning*, 37(1), 285–296. <https://doi.org/10.1111/j.1467-1770.1987.tb00>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.
- Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, 24(2), 125–146. <https://doi.org/10.1177/002383098102400202>
- Ghulamullah, & Hamzah, M. H. (2017). Intelligibility and comprehensibility issues among Pakistani speakers of English. *International Journal of English Language, Literature and Translation Studies*, 4(2), 506–513. <http://www.ijelr.in/4.2.17a/506-513%20GHULAMULLAH.pdf>
- Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, 2(1), 56–92. <https://doi.org/10.1075/jslp.2.1.03gor>
- Hamzah, M. H. (2010). Durational properties of initial geminate consonants in Kelantan Malay. In Tabain, M., Fletcher, J., Grayden, D., Hajek, J., & Butcher, A. (Eds.), *Proceedings of the 13th Australasian international conference on speech science and technology* (pp. 18–21). ASSTA.
- Hamzah, M. H., Ahmad, A., & Yusuf, M. H. (2017). A comparative study of pronunciation among Chinese learners of English from Malaysia and China: The case of voiceless dental fricatives /θ/ and alveolar liquids /r/. *Sains Humanika*, 9(1), 1–9. <https://doi.org/10.11113/sh.v9n4-2.1354>
- Hamzah, M. H., & Bawodood, A. M. A. (2019). Teaching English sounds via minimal pairs: The case of Yemeni EFL learners. *Veda's Journal of English Language and Literature*,

- 6(3), 97–102.
<https://doi.org/10.33329/joell.63.97>
- Hamzah, M. H., Fletcher, J., & Hajek, J. (2011). Durational correlates of word-initial voiceless geminate stops: The case of Kelantan Malay. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th international congress of phonetic sciences* (pp. 815–818). University of Hong Kong.
- Hamzah, M. H., Fletcher, J., & Hajek, J. (2016). Closure duration as an acoustic correlate of the word-initial singleton/geminate consonant contrast in Kelantan Malay. *Journal of Phonetics*, 58(1), 135–151.
<https://doi.org/10.1016/j.wocn.2016.08.002>
- Hamzah, M. H., Hajek, J., & Fletcher, J. (2020). Non-durational acoustic correlates of word-initial consonant gemination in Kelantan Malay: The potential roles of amplitude and f0. *Journal of the International Phonetic Association*, 50(1), 23–60.
<https://doi.org/10.1017/S0025100318000142>
- Huwari, I. F., & Mehawesh, M. (2015). Review of the importance of teaching pronunciation in the Arab society. *International Journal on Studies in English Language and Literature*, 3(6), 31–37.
- Ismael, A., Mahadin, D., & Masri, A. A. (2015). Difficulties facing English language students at Al Balqa Applied University in English pronunciation. *Global Journal of Art and Social Science Education*, 3(1), 72–81.
- Kager, R., van der Feest, S., Fikkert, P., Kerkhoff, A., & Zamuner, T. S. (2007). Representations of [voice]: Evidence from acquisition. In J. van de Weijer & E. J. van der Torre (Eds.), *Voicing in Dutch: (De)voicing – phonology, phonetics, and psycholinguistics* (pp. 41–80). John Benjamins.
- Khattab, G. (2000). VOT production in English and Arabic bilingual and monolingual children. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*(8), 95–122.
<https://doi.org/10.1075/cilt.230.03kha>
- Khattab, G. (2002). VOT in English and Arabic bilingual and monolingual children. In D. Parkinson & E. Benmamoun (Eds.), *Perspectives on Arabic linguistics XIII-XIV: Papers from the thirteenth and fourteenth annual symposia on Arabic linguistics* (pp. 1–38). John Benjamins Publishing Co.
- Ladefoged, P., & Maddieson, I. (1998). The sounds of the world's languages. *Language*, 74(2), 374–376. <https://doi.org/10.2307/417875>
- Lee, H., & Hwang, H. (2016). Gradient of learnability in teaching English pronunciation to Korean learners. *The Journal of the Acoustical Society of America*, 139(4), 1859–1872. <https://doi.org/10.1121/1.4945716>
- Lerdpaisalwong, S. (2015). *Perception training of Thai learners: American English consonants and vowels* [Unpublished doctoral dissertation]. The University Of Wisconsin-Milwaukee. <https://dc.uwm.edu/etd/1009>
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Words*, 20(1), 384–422.
<https://doi.org/10.1080/00437956.1964.116598>
- Lisker, L., & Abramson, A. S. (1967). Some effects of context on Voice Onset Time in English stops. *Language and Speech*, 10(1), 1–28.
<https://doi.org/10.1177/002383096701000101>
- Mahmood, R. M. (2016). A contrastive study of the Voice Onset Time (VOT) in English and Arabic languages. *Arts Journal*, 118(1), 61–74.
<https://doi.org/10.31973/aj.v1i118.374>
- Naser, I. M. M., & Hamzah, M. H. (2018). Pronunciation and conversation challenges among Saudi EFL students. *Journal of English Educators Society*, 3(1), 85–104.
<https://doi.org/10.21070/jees.v3i1.1228>
- Olson, D. E. (2017). *Voice Onset Time in Arabic and English stop consonants* [Unpublished bachelor's thesis]. The University of Utah. <https://speechlab.utah.edu/2017%20Olson%20Thesis.pdf>
- Olson, D. E. & Hayes-Harb, R. (2019). Voice onset time in Arabic and English stop consonants. *Al-Arabiyya*, 52(1), 29–48.
- Port, R. F., & Mitleb, F. M. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics*, 11(1), 219–229.
[https://doi.org/10.1016/S0095-4470\(19\)30823-X](https://doi.org/10.1016/S0095-4470(19)30823-X)
- Rato, A., & Rauber, A. S. (Eds.). (2015). *Proceedings of the 18th international congress of phonetic sciences*. Glasgow University.
- Sabbah, S. (2015). Negative transfer: Arabic language interference to English learning. *Arab World English Journal*, 4(1), 269–288.
<https://doi.org/10.2139/ssrn.2844015>
- Shormani, M. Q., & Alsohbani, Y. A. (2015). Phonological transfer and universal grammar: Evidence from Arabic. *Journal of Teaching and Teacher Education*, 3(2), 37–48.
<https://doi.org/10.12785/jtte/030203>
- Simon, E. (2009). Acquiring a new second language contrast: An analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research*, 25(1), 377–408.
<https://doi.org/10.1177/0267658309104580>
- Yeni-Komshian, G. H., Caramazza, A., & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics*, 5(1), 35–48.
[https://doi.org/10.1016/S0095-4470\(19\)31112-X](https://doi.org/10.1016/S0095-4470(19)31112-X)