Park or bark? VoT Production of English Bilabial Stops in Arabic Learners of English

XXX

XXX
XXX

ABSTRACT

This preliminary study compares VoT patterns of word-initial English /b/ and /p/ in three groups: Omani undergraduate students, Omani professors and English native speakers (ENS). It also compares VoT patterns of these sounds in two phonetic environments: CV and CCV. Words containing initial /b, p/ were elicited using a reading task. The results revealed a significant difference in VoT patterns of /p/ between students and ENSs, but not between professors and ENSs. Students' /p/ was characterized with shorter lag compared to professors' or ENSs'. Variability in the production of /b/ existed among Omani speakers. While ENSs' VoT productions were characterized with short lag, students' and professors' productions were characterized sometimes with voicing lead and others with short/long lag. VoT values of /b/ and /p/ were not significantly affected by whether the sound was in a CV or CCV syllable. The findings are discussed in relation to speech learning research.

Keywords: VoT, voicing lead, voicing lag, speech learning

1. INTRODUCTION

There has been a growing number of studies examining the acquisition of English stops by Arabic native speakers [1, 2, 3, 4, 5]. The main acoustic dimension used to distinguish Arabic-accented stops from English stops in these studies is voice onset time (VoT). Lisker and Abramson [6, p.422] defined VOT as "the time interval between the burst that marks release and the onset of periodicity that reflects laryngeal vibration". VoT is found to be highly reliable in separating homorganic stop categories despite phonetic differences among languages [6].

The VoT patterns of English and Arabic word-initial stops vary considerably. English /p, t, k/ are normally aspirated while /b, d, g/ are unaspirated. Aspirated English stops are produced in the long-lag region and have positive VoT values. Glottal pulsing, which distinguishes voiced and voiceless sounds, during closure is not necessarily available in the production of English stops. Hence, English word-initial /b, d, g/ are produced in the short-lag region. On the other hand, Arabic /b, d, g/ are normally accompanied by glottal pulsing during closure (i.e. voicing lead) and thus are at the opposite end of the continuum compared to English voiceless aspirated stops [5]. The VoT range of Arabic voiceless stops falls within the range of English unaspirated stops /b, d, g/ [5].

Since Arabic stops, /b, d, t, k/, overlap with those in English, English speakers face no difficulty in accurately identifying them when Arabic learners of English attempt to speak English [1]. However, it is widely acknowledged that Arabs have a persistent problem in the production of acceptable English /p/ due to the absence of this sound from the Arabic phonemic inventory [1, 3, 4]. Arabs tend to either substitute English /p/ with Arabic /b/ or produce it with short lag [3, 4].

The Contrastive Analysis Hypothesis (CAH) assumes that L2 phonemes that have no equivalent counterpart in the L1 would be difficult to learn while those that have an L1 equivalent would be relatively easy to learn [7]. However, Flege [8] argues that the CAH ignored the fact that two 'similar' sounds in two languages may differ vastly at the phonetic level. The Speech Learning Model (SLM) and more recently SLM-r predicts that sounds that are 'similar' in two languages may be harder to learn than sounds that are 'new' [8]. It also proposes that L2 learners gradually discriminate between L1 and L2 phonetic details as they gain more experience using the L2 in everyday life [9]. Based on this, one can predict that Arabs, with sufficient input, will eventually produce a satisfactory English /p/. Word-initial English /b/, on the other hand, might be assimilated to Arabic /b/ and produced with glottal pulsing during the closure period.

Although there has been a rise in the number of instrumental studies interested in examining the acquisition of the phonological/phonetic features of English stops by native Arabic speakers, this area of research remains understudied, especially research that examines Omani Arabic speakers of English. Additionally, most of these studies have looked at

whether native Arabic speakers have acquired the distinction between voiced and voiceless stops in a CV syllable. This study is interested in extending this research to include an examination of the VoT patterns of these stops in a consonant cluster. We predict that Omani speakers of English might find it harder to produce native-like VoT when the stop is in CCV syllable than when it is in a CV syllable. Thus, our study aims to: 1) compare VoT production of English /b/ and /p/ in three groups of speakers: Omani undergraduate students, Omani professors and English native speakers (ENS), 2) examine the effect of the phonetic environment on VoT production.

2. METHOD

2.1. Participants

Three groups of participants were recruited for this study. The Omani Native Arabic speakers fall into two groups. The first group involved nine male and female students who had studied English in Omani public schools from grade one through grade 12, and were specializing in English language at the university level (BA) at the time of the study. All were 20-22 years old and had finished at least 30 credits of English language and content courses. The second group of Omani native speakers involved 6 instructors with PhD in Linguistics or English Language that they obtained from the USA, Canada or the UK. This means that they spent at least 7 years residing in a native English country. All were 37-45 years old and taught English specialized courses at the university BA level. Four native English speakers (British, Australian and American) who were between 33 and 47 years old participated as a control group.

2.2. Material

31 English monosyllabic words containing word-initial /b/ and /p/ were used in this study. The words were randomized in a typed list and embedded in the carrier sentence "I like to say…". 16 of the words contained /p/, six of which contained /p/ preceding a vowel as in *pack, pig, path, peek, page, pad*, while 10 of which contained /p/ in a consonant cluster, preceding a sonorant consonant /r/ or /l/ as in *plank, plate, play, place, plane, prop, prat, prig, pram, proud*. 15 of the words contained /b/, 5 of which contained /b/ before a vowel as in *big, bad, beat, bath, back*, 10 words contained /b/ before a sonorant consonant as in *blue, black, blood, bleed, block, broad, brag, brig, broom, bread* and *break.*

2.3. Procedure

The participants were instructed to read the sentences at a comfortable tempo. They were recorded on Praat, which was digitized at 44.1 kHz, using Microsoft headphones and a microphone in a quiet room at the English Department at Sultan Qaboos University. One British English speaker recorded herself on Praat using a microphone in her home in the UK.

2.4. Acoustic and statistical analyses

526 words were submitted to wideband spectrograms in Praat. For each test word, two temporal intervals were measured by hand to the nearest 5 ms from the spectrogram. VoT of /p/ was measured from the onset of the release burst to the point where the first periodic cycle starts. VoT of pre-voiced productions were measured from the offset of the preceding vowel (indicated by a drop in energy in the region of F1 and F2) to the onset of release burst. A binary judgement for the presence or absence of glottal pulsing during the closure interval of speakers was made [1]. A stop was considered to have been produced with glottal pulsing (pre-voicing) if it showed visible periodic striation throughout most of the closure [1, 10].

Statistical analyses were carried out in R [11]. Linear Mixed Effect Models (LMER) were used to generate statistical evidence for the differences in VoT values among the speakers using lme4 package [12]. For fixed effects, we used group (ENS, student, professor), the following sound and an interaction between group and following sound. For random effects, we used speaker and word as random intercepts.

3. RESULTS

3.1 VoT patterns of /p/

In some students and professors, the production of /p/ was accompanied with glottal pulsing. There were two instances of glottal pulsing in the production of /p/ in professors and 7 instances of glottal pulsing in students' productions. This indicates that English /p/ was assimilated to Arabic /b/ in these productions. These tokens were removed from the rest of the analyses.

LMER revealed a significant difference in mean VoT between students and ENSs ($p<0.01$). ENSs' mean VoT was 54.29 ms ±10.87 (standard error) higher than that of students'. The difference in mean VoT of /p/ between professors and ENSs was not significant ($p>0.05$). ENSs' mean VoT was only 17.02 ms ±11.59 (standard error) higher than that of professors. This indicates that professors' mean VoT of /p/ falls within the range of ENSs'. The mean VoT

of professors was also significantly different from students' ($p<0.01$). Professors' mean VoT was 37.26 ms ±9.87(standard error) higher than that of students. As to the effect of the phonetic environment on VoT patterns of /p/, LMER revealed no significant difference between the mean VoT of /p/ before a vowel and the mean VoT of /p/ before a consonant. Visual inspection of Figure 1 shows that in all three groups of speakers, the mean VoT of /p/ is higher when followed by a sonorant consonant than when followed by a vowel, though this was not significant.

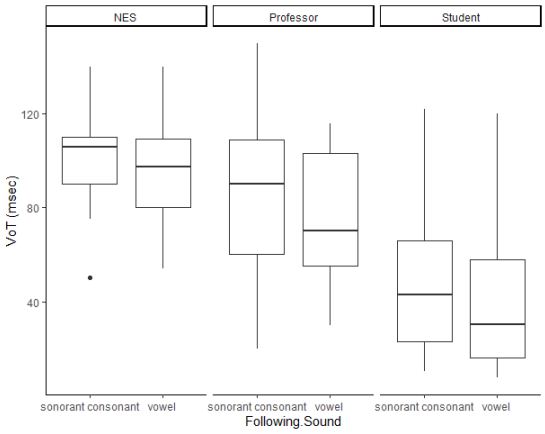

Figure 1. VoT values of /p/ in three groups of speakers distributed based on following sound

It is worth noting that variability in the VoT values of /p/ existed among Omani native speakers. Table 1 shows that the range of VoT values in students' productions is between 7 ms and 122 ms and that in professors is between 20 ms and 150 ms. The minimum VoT value in ENSs is 50 ms. This indicates that some +VoT productions in Omani speakers were within the short lag range, which characterizes English /b/.

	Min	Mean	Max
ENS	50	99	140
Student	7	43	122
Professor	20	81	150

Table 1. Minimum, maximum and mean VoT values of /p/ in three groups of speakers

3.2. VoT patterns of /b/

When measuring VoT productions of /b/, LMER revealed a significant difference between professors and ENS' ($p<0.01$). Professors' mean VoT was 70.65 ms ±22(standard error) lower than that of ENSs'. On the other hand, the mean VoT in students was not statistically different from that of ENSs' ($p>0.05$). These findings might be misleading if we ignored the variability in VoT productions in native Omani speakers. LMER also revealed no significant effect of the following sound on VoT productions of /b/.

The VoT patterns of /b/ varied within and across the three groups (Table 1). While ENSs' productions were mostly characterized with short lag and no voicing lead, students' productions were characterised with voicing lead (20%), short and long lag (80%). Professors' productions were characterised mostly with voicing lead (70%) and to a lesser extent short lag (29%).

/b/	N (%)	Group	Min	Mean	Max
-VoT	50 (20)	Student	-150	-98	-50
	71 (70)	Professor	-130	-97	-60
	59 (100)	ENS	5	14	40
+VoT	116 (80)	Student	8	33	127
	26 (29)	Professor	10	19	40

Table 2. Mean, minimum and maximum VoT values of /b/ in three groups of speakers

When measuring the difference in mean +VoT between the three groups, LMER revealed that ENSs' mean VoT was 21.06 ±7.68 (standard error) lower than that of students ($p<0.01$). On the other hand, ENSs mean VoT was not significantly different from mean VoT of professors ($p>0.05$). This indicates that students' +VoT of /b/ was exaggerated compared to ENSs'. From Table 1, we can notice that the maximum VoT value of /b/ in students is 127 ms, a value that falls within the VoT range of /p/. With regard to the effect of the following sound on +VoT values of /b/, LMER revealed no significant effect of the following sound on mean +VoT within the three groups.

4. DISCUSSION

This study shed light on the VoT patterns of English /b/ and /p/ in the productions of three groups of speakers: Omani undergraduate students, Omani professors and ENSs. It also examined any effects of the phonetic environment in which /b/ and p/ appear.

The study found that Omani professors' VoT productions of /p/ were not different from ENSs'. As ENSs, professors produced their /p/ with long lag (i.e. aspiration) most of the time. This finding is not in line with previous studies [1, 2, 3, 4] that found Arabs'

VoT productions of /p/ fall mostly within the short lag region (below 40 ms). Similar to previous research, however, students' VoT productions were significantly different from those of ENSs and professors'. Mean VoT value of students in all phonetic environments (m= 43 ms) was not near that of ENSs' (m= 99 ms) or professors' (m= 81 ms).

One justification for the native-like production of /p/ in professors is the amount of authentic exposure they have had in the target language. Professors' success in the production of a native-like /p/ compared to students may be in part due to their experience in using English [13]. These speakers spent 7 or more years studying English and linguistics in a native speaking country. Undoubtedly, they might have formed a new phonetic category for /p/ in their phonetic inventory [8, 14]. Hence, this finding does not provide support to the CAH which predicts that L2 learners will face difficulty acquiring sounds that are different from those in their phonemic inventory. Another justification is the type of task the speakers were asked to perform to elicit data for this study. L2 learners are found to perform better in reading tasks than in spontaneous speech because they have more time to access their explicit knowledge [15]. It would be interesting to examine native Omani speakers' productions of /p/ using more naturalistic data-elicitation tasks.

Despite the fact that professors' productions of /p/ were native-like in terms of VoT, it is worth noting that there was also variability in the range of VoT values in professors' productions. Some VoT productions fell within the short lag region and others fell within the long lead region. This points to the importance of considering individual differences and even differences within speakers when making conclusions about L2 acquisition.

The study also found that there was large variability in the VoT productions of /b/ especially among students. Professors' and students' productions were sometimes accompanied with voicing lead and other times short lag but also long lag in students. ENSs' productions of /b/, on the other hand, were accompanied with short lag. This is in line with research that confirmed the absence of prevoicing in word-initial /b/ of ENSs [1, 10]. Overall, the results of the current study showed that the mean VoT of professors' /b/ was significantly lower than that of ENSs' or students'. This is obviously because professors' /b/ was mostly accompanied with voicing lead. Interestingly, this supports the SLM which predicts that L2 learners will find it harder to acquire a sound that is phonetically similar to another sound in their phonetic inventory [8, 9].

Does the previous finding suggest that students were more successful at producing a native-like /b/? It is hard to answer this question with yes just by considering the previous finding. Further analyses of +VoT productions of /b/ revealed that students' mean VoT (m= 33 ms) was significantly higher than that of ENSs' (m=14 ms). Some students' +VoT productions fell within the VoT range of /p/. This might be one reason why students' VoT productions of /b/ were statistically similar to ENSs'. In a study that investigated Saudi learners' acquisition of English stops, Alanazi [4] found that his Arab speakers also produced English /b/ with glottal pulsing, short lag and long lag. Apparently, producing English /b/ with long lead is an influence of the speakers' L1, Arabic [16, 17]. On the other hand, producing /b/ with long lag might be due to the pressure of the reading task the learners had to perform. The learners, being at a developmental stage, probably got confused as to whether they should pronounce /p/ or /b/ as the reading list contained randomized words with either sound word-initially.

The phonetic environment of /p/ and /b/ (CV or CCV) did not have a significant effect on VoT productions in any of the three groups of speakers. The VoT patterns of /p/ preceding a vowel were lower than those of /p/ in a consonant cluster, though this was not significant. This finding is in line with a study by Klatt [18] who found that the VoT of word-initial English voiceless stops was longer before sonorants and high vowels than before mid and low vowels.

Taken all together, the findings of this study confirm that experienced L2 learners are capable of producing native-like sounds that are novel especially when the data-elicitation task stimulates the learners' explicit knowledge. It also suggests that L2 learners may not be able to acquire some phonetic details of L2 sounds that are perceived as similar to L1 sounds. The study calls for the importance of considering variability in L2 productions. The L2 speakers in this study showed large variability in the range of VoT productions of their English /p/ and /b/.

5. REFERENCES

[1] Flege, J. E., & Port, R. 1981. Cross-language phonetic interference: Arabic to English. *Language and speech*, *24*(2), 125-146.

[2] Port, R. F., & Mitleb, F. M. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. *Journal of Phonetics*, *11*(3), 219-229.

[3] Sawsan, A. 2017. An acoustic analysis of the production of word-initial stop/p/by late Arab bilinguals. *Training, Language and Culture*, *1*(4), 68-83.

[4] Alanazi, S. 2018. *The acquisition of English*

stops by Saudi L2 learners (Doctoral dissertation, University of Essex).
[5] Khattab, G. 2002. VOT production in English and Arabic bilingual and monolingual children. *Amsterdam Studies in the Theory and History of Linguistic Sciences Series* 4, 1-38.
[6] Lisker, L and Abramson A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. WORD 20(3), 384-422.
[67] Lado, R. 1964. Linguistics across cultures, University of Michigan press. *Ann Arober, Michingan*.
[8] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233-277.
[9] Flege, J. E., & Bohn, O. S. 2021. The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, 3-83.
[10] Abramson, A. S., & Whalen, D. H. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of phonetics*, *63*, 75-86.
[11] Team, R. C. 2013. R: A language and environment for statistical computing.
[12] Bates D., Mächler M., Bolker B. & Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1–48. doi:10.18637/jss.v067.i01.
[13] Piske, T., MacKay, I. R., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, *29*(2), 191-215.
[14] Flege, J. E., Schirru, C., & MacKay, I. R. 2003. Interaction between the native and second language phonetic subsystems. *Speech communication*, *40*(4), 467-491.
[15] Saito, K., & Brajot, F. X. 2013. Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English/ɹ/by late JapaENSe bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 847-863.
[16] Ringbom, H. 1992. On L1 transfer in L2 comprehension and L2 production. *Language learning*, *42*(1), 85-112.
[17] Smith, M. S., & Kellerman, E. 1986. Crosslinguistic influence in second language acquisition: An introduction. *Crosslinguistic Influence in Second Language Acquisition, New York: Pergamon*.
[18] Klatt, D. H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of speech and hearing research*, *18*(4), 686-706.