Does Learning English Vowels Aid the Perception of English Stress for Learners with Small Vowel Inventory Background?

Manal Allehyani, Rachel Smith, & Clara Cohen
2503139a@student.gla.ac.uk
Glasgow University Laboratory of Phonetics, University of Glasgow

ABSTRACT

This study is part of a larger longitudinal project investigating the relationship between acquiring English vowel quality and acquiring English stress by Arabic learners of English. In this paper, two experiments were conducted to test 1) learners' ability in perceiving vowel contrasts of different difficulty levels, and 2) the relationship between the ability identify vowels and the ability to perceive stress placement in disyllabic nonsense words. An analysis using Signal Detection Theory revealed that all dialect groups were less sensitive to the quality of vowels in the difficult pairs---i.e., pairs distinguished by contrasts absent in the native Arabic dialects---relative to medium and simple pairs. Further, the participants' d-prime scores in vowel identification and stress placement were positively correlated. These findings suggest that acquisition of English stress patterns may benefit from learners' ability to perceive vowel contrasts.

Keywords: stress perception, vowel perception, signal detection theory, language acquisition, Arabic

1. INTRODUCTION

It is well-established in the field of second language acquisition (SLA) that the first language (L1) influences the way second language (L2) learners perceive the target language [1-2]. Arabic and English differ in their vowel inventories, in the phonetic cues used to signal stress, and their phonological stress patterns. Each of these differences makes perceiving and producing English stress a challenging task for Arabs.

Languages' vowel inventories vary drastically in size. Modern Standard Arabic has a small inventory of 3 short vowels /a/, /ɪ/, /u/ and their long counterparts /a:/, /ɪ:/, and /u:/. Some regional variants contain other vowels, such as /e:/ and /o:/, or schwa in Moroccan and Syrian dialects [3]. English, by contrast, has a large inventory of twelve monophthongs and eight diphthongs. This makes it challenging for beginner Arab learners to discriminate English contrastive vowels: learners tend to assimilate several contrastive L2 vowels to the same L1 [4]. Most Arabic dialects also lack central vowels, leading learners to assimilate English central vowels to high or low Arabic categories. This not only hinders their perception and production of vowel contrasts but also creates difficulties in perceiving and producing English stress, which exploits vowel quality as a segmental cue to stress.

The two languages' differences in vowel inventory have implications for the phonetic parameters they use to maintain perceptual contrast between vowels. In terms of vowels' intrinsic F0 (IF0), English speakers maximise the perceptual difference between the F0 of high and low vowels to free more perceptual space for half-open vowels. This is reported to enhance vowel contrasts [5] which is imperative since the crowded vowel space causes some vowels' ellipses to overlap [6]. However, the scenario is different in Arabic. The very small vowel inventory and the lack of many half-open vowels encourage the perceptual difference of IF0 between high and low vowels to be kept to the minimum since the vowel space is not as crowded as it is in English. However, Arabic vowels differ mainly in the dimension of height, with little variation in IF0. Arabic speakers rely more on lengths to maximize vowel contrasts [7,3]. Among the Arabic dialects there are variation in the duration and ratio between long/short vowels. Long and short vowels in Sudanese dialect were reported as longer than their long and short counterparts in Jordanian and Libyan dialects. The Libyan dialect have the shortest vowels durations and the smallest short to long ratios whereas the Jordanian dialect lies in between [8]. In addition, Arabic vowels undergo a process of velarization after velarized consonants. In such vowels, the tongue is lowered and retracted, raising F1 and lowering F2, and resulting in higher intensity values compared to the plain vowels.

Intensity and f0, along with duration, are the major phonetic cues to stress [9]. This brings us to theorize that if IF0 values of the vowels are necessary to enhance vowel contrasts in English, English speakers should use F0 and vowel quality to signal stress because those cues also serve to enhance vowel contrasts. However, since Arabic does not link stress and vowel quality in this way, then Arabic speakers may rely more on their native cues that do not signal vowel contrasts in English—namely, intensity and length. Indeed, studies have shown that the perceptual weighting of stress cues is language-specific [10] in ways that are broadly consistent with this hypothesis.

In English, vowel quality is weighted as the strongest cue to stress (i.e., where a stressed syllable triggers vowel reduction in an adjacent unstressed syllable [11-13]) followed by pitch [14,15], which is in turn more important than duration and intensity [16-19]. In Arabic, by contrast, vowel quality does not play a role in perceiving or producing stress [20-25]. Instead, Arabic speakers rely more on using higher amplitudes and longer vowel durations when producing contrastive stress in Arabic. When learning English, Arabs produced higher F0 than native English speakers which indicates that they are aware of the need to use this cue to highlight stress in producing English stressed syllables.

Regarding stress placement, English is considered as a non-predictable stress language which does not rely on syllable weight when assigning stress. On the other hand, Arabic is classified as a highly predictable-stress language whose stress assignment rules are sensitive to syllable weight. A superheavy ultimate (i.e., CVVC or CVCC) receives stress, otherwise the penultimate receives stress only if heavy (i.e., CVC or CVV) [26]. For this reason, English stress is partially encoded in the lexicon whereas Arabic stress is assigned post-lexically. Moreover, since Arabic stress is predictable, Arabs struggle in Arabic stress perception tasks, producing lower accuracy and higher reaction times than English and Chinese speakers [27]. This deafness of Arabic native speakers may be a by-product of predictability.

We argue that there is a close connection between the factors that make English vowels hard to discriminate for Arabic learners, and the factors that complicate English stress perception for them. The main hypothesis in this paper is thus that learning English vowels will help native speakers of Arabic to better perceive English stress. As a preliminary test of this hypothesis, two perceptual experiments were administered: a vowel perception experiment and a stress placement perception experiment. If improving English vowel perception helps listeners with English stress perception, then learners who are better at perceiving English vowel contrasts should also show better stress perception than learners who struggle to hear those contrasts. It is important to note that work is in progress to gauge native Arabic speakers' sensitivity to each of the stress cues (i.e., intensity, vowel length, and pitch), and how their performance in each of these experiments is correlated to their performance in the perception of English vowels. After six months of exposure to English, the same participants will reconduct the same experiments to see if causation can be established between their vowel perception and their perception in the other three experiments of 1) stress placement perception, 2) perception of pitch experiment, 3) perception of vowel length, and 4) perception of intensity in the two phases of the project. Here, we produce the preliminary results of this project, demonstrating a strong connection between vowel perception and stress placement perception in exactly the direction that would be expected to aid listeners in acquiring English vowel contrasts.

2. METHODS

2.1. Vowel perception experiment

2.1.1. Participants

40 Jordanians, 40 Libyans and 39 Sudanese (mean ages of 33.7, 33.4, and 30.5, respectively) participated in the two experiments in this paper and all other experiments of the larger project. They all live in England.

2.1.2. Materials and procedure

The participants were tested on their ability to discriminate 18 contrastive vowel pairs in an AXB task. Table 1 shows pairs varied in difficulty level (i.e., high, medium, low, plus four easy pairs used as distractors). Difficulty levels were determined based on 1) frequency of reporting problematic English pairs for speakers from different Arabic backgrounds in the TESOL literature [28] and 2) results of perceptual assimilation tasks and vowel confusion matrices in phonetic studies [29, 30].

Pair (A-B)	Target (A)	Difficulty
BATH-TRAP	BATH	High
DRESS-KIT	DRESS	High
LOT-STRUT	LOT	High
LOT-FOOT	LOT	High
STRUT-TRAP	STRUT	High
STRUT-BATH	STRUT	High
LOT-CAUGHT	LOT	Medium
NURSE-FLEECE	NURSE	Medium
CAUGHT-NURSE	CAUGHT	Medium
DRESS-STRUT	DRESS	Medium
LOT-BATH	LOT	Medium
STRUT-FOOT	STRUT	Medium
CAUGHT-GOOSE	CAUGHT	Low
NURSE-STRUT	NURSE	Low
NURSE-BATH	NURSE	Low
GOOSE-FOOT	GOOSE	Low
KIT-FLEECE	KIT	Low
CAUGHT-BATH	CAUGHT	Low

Table 1: The contrastive pairs, their targets and difficulty level.

The A and B stimuli were of the form "bVd" and the X options were of the form "hVd". Each pair was manipulated as AAB, ABB, BAA, and BBA. For example, the pair DRESS-KIT was manipulated as DRESS-DRESS-KIT, DRESS-KIT-KIT, KIT-DRESS-DRESS, and KIT-KIT-DRESS. The X options were equally either A (e.g., DRESS) or B (i.e., KIT). Each manipulation was repeated twice yielding 8 repetitions per contrastive pair (i.e., 21 contrastive pairs * 8 repetitions = 168 responses/participant). Using Gorilla, the participants were asked to press a key to indicate whether X resembled the first or second word. The experiment lasted for 45 minutes.

2.1.3. Analysis

A signal detection analysis was conducted. The target was always defined as the L2 phoneme that, according to the literature, the participants should find more difficult to learn. This difficulty may be because the phoneme is perceptually new, or because, if it exists in L1, it must be licensed by a specific phonological context. Note that the phoneme difficulty level is relative, based on the other phonemic member within the contrastive pair. For example, the vowel of BATH was the target for the BATH-TRAP pair, but not for the STRUT-BATH pair. In this sense, (A) of each pair was specified as the target (Table 1). AAB and BAA sets were specified as the target groups (i.e., X=A in AAB and BAA) whereas ABB and BBA as the nontarget group (i.e., X=B). Hits, misses, false alarms (FA), and correct rejections (CR) were thus coded as in Table 2.

Stimuli	**Response**	
	X=A	X=B
Target (X =A) in AAB and BAA stimuli	Hit	Miss
Non-target (X=B) in BBA and ABB) stimuli	FA	CR

Table 2: Vowel experiment coding for signal detection analysis.

It was hypothesized that all learner groups would perform worst for high difficulty pairs, represented by lower d-primes suggesting that the pairs are still perceived as the same L1 category, Medium and low difficulty pairs should produce higher d-prime scores. We also hypothesised that the Sudanese group would perform the worst since they have the largest length difference between long and short vowels. The Libyans were hypothesized to perform better than the Sudanese for having the smallest ratio of long and short vowel durations.

2.2. Stress placement experiment

2.2.1. Materials and procedure

The stimuli in this task are disyllabic nonsense words of the form CVC.CVC and CVC.CV:C (e.g., [tɪn.zɪn] and [nɪm.dɪ:f], respectively). The main vowel was in the second syllable of each word form; the 11 monophthongs shown in Table 1 each formed the main vowel in nine different stimuli, with a high, mid and low first-syllable vowel respectively. For each stimulus, stress either fell on the first or second syllables. This resulted in 198 stimuli/participant (i.e., 11 vowels * 9 stimuli/vowel * 2 stress locations). The participants listened to each stimulus once and were asked to press a key to indicate whether stress fell on the first or second syllable. The whole task lasted for a maximum of thirty minutes.

2.2.2. Analysis

A signal detection analysis was again conducted. Based on Arabic’s predictable rules, stimuli of the form CVC.CVC attract stress to their first syllable and CVC.CV:C forms attract stress to their second syllable. If the participants were answering based on these Arabic rules, this would indicate that they were insensitive to the phonetic cues to stress in English. Thus, the target for the performance matrices was defined as a response that showed a correct ability to ignore the phonological rules of Arabic stress. This means in CVC.CVC forms, the target is the second stressed syllable and in CVC.CV:C forms, the target is the first stressed syllable. The two targets were combined to generate one d-value per participant (Table 3).

Stimuli	**Response**	
	Target	Non-target
Targets (**CVC**.CV:C and CVC.**CVC**)	Hit	Miss
Non-targets (CVC.**CV:C** and **CVC**.CVC)	FA	CR

Table 3: The targets in CVC.CVC and CVC.CV:C forms combined. Bold font indicates stress.

We hypothesised that the participants who could successfully discriminate vowels in the vowel perception experiment would perform well in this experiment too, showing a positive correlation between acquiring English vowels and acquiring English-like cues to stress. This will be evaluated by testing the strength and direction of the correlation between the d-primes of the two tasks for each participant.

3. RESULTS

3.1. Vowel perception results

Participants’ d-prime scores were calculated for each difficulty level per pair, producing three d-prime scores per participant. These scores were then analysed with linear mixed effects regression modelling, using the `lme4` package [31] in `R` [32] P-values for coefficient estimates were calculated based on Satterthwaite denominator degrees of freedom, as provided by the `lmerTest` package [33]. D-prime was modelled as a function of Difficulty Level, Dialect, and their interaction as fixed effects, with a random intercept included for Participant.

The model summary is provided in Table 4, with a partial effects plot illustrating the patterns in Figure 1. There was a significant interaction between Difficulty Level and Dialect. For Jordanian speakers, d-prime was significantly higher with low and medium difficulty pairs than for high difficulty pairs (low $\beta = 0.93$, $p < .001$; med β=0.76, p < .001). Libyan speakers did not differ significantly from Jordanian speakers in this respect, and for medium difficulty pairs, Sudanese speakers also did not differ from Jordanians. For low-difficulty pairs, however, Sudanese speakers showed less improvement than Jordanians (β = -0.64, $p < .001$).

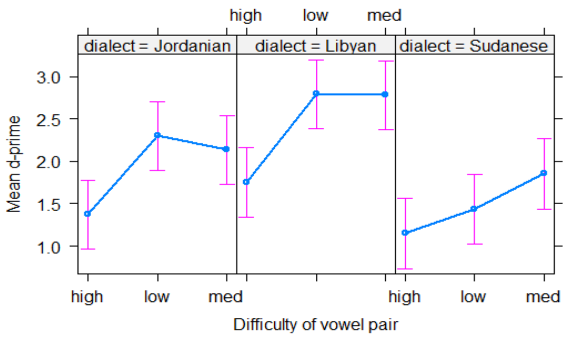

Figure 1: Partial effects plot showing model effects of Difficulty (low, med, high) and Dialect (Jordanian, Libyan, Sudanese) on vowel perception D-prime scores.

Effect	β	SE(β)	t	
Intercept	1.37	0.21	6.64	***
Dif=low	0.93	0.12	7.48	***
Dif=med	0.76	0.12	6.16	***
Dia=Lib	0.38	0.29	1.30	
Dia=Sud	-0.23	0.29	-0.77	
Dif=low × Dia=Lib	0.11	0.17	0.66	
Dif=med × Dia = Lib	0.27	0.17	1.52	
Dif=low × Dia=Sud	-0.64	0.18	-3.65	***
Dif=med × Dia = Sud	-0.06	0.18	-0.33	

Table 4: Summary of mixed effects model predicting vowel perception d-prime against Difficulty (low, med, high) and Dialect (Jordanian, Libyan, Sudanese). All factors were treatment-coded. For Dialect, Jordanian was the reference level; for Difficulty, High was the reference level. Three asterisks indicate $p < .001$.

3.2. Stress placement perception results

To examine the relationship between vowel and stress-placement perception, D-prime scores were calculated on the stress placement task, and re-calculated for the vowel perception task, this time collapsing across all difficulty levels. This produced one d-prime score per listener for each task. As illustrated in Figure 2, the d-prime scores showed a significant positive correlation across tasks (ρ = .363, t(116)=4.20, $p < .001$). The R^2 of 0.13 indicates that about 13% of the variance in stress placement d-prime was explained by variance in vowel perception d-prime.

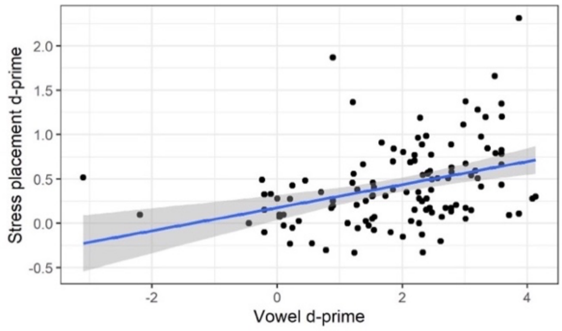

Figure 2: Correlation between vowel perception D-prime scores and stress placement D-prime scores.

4. DISCUSSION AND CONCLUSION

This paper argued for the possible role that learning English vowel contrasts might play, for Arab learners of English, in perceiving English stress location in contrastive disyllabic words. We showed that Arabic learners do struggle to discriminate vowel pairs that prior literature has identified as challenging. Speakers of Jordanian, Libyan and Sudanese Arabic pattern similarly to each other, except that Sudanese learners struggle slightly more with easy contrasts. Crucially, across the two AXB tasks we observed the predicted significant positive correlation between perceiving English vowels and perceiving English stress placement.

We think this relationship is best explained in terms of the two languages' differences in vowel inventories. More specifically, we speculate that the secondary cues that a language uses to differentiate the vowels in its inventory, will be the same as those used to signal stress. The condensed vowel space of English necessitates maximising the differences of IF0 values between high and low vowels to create more space for half-open vowels. In Arabic, by contrast, as a small vowel inventory language with only two half-open vowels (i.e., /e:/ and /o:/), no maximization of the difference in IF0s is required since the vowel space is not crowded. In addition, Arabic vowels are differentiated by length, and pharyngealization processes contribute to intensity differences. We therefore hypothesize that when applying stress, the two languages use different cues that distinguish their vowels. English uses vowel quality and pitch whereas Arabic uses duration and intensity. Work is in progress to test this relation by assessing Arabic learners' ability to discriminate subtle variation in the three stress cues of pitch, intensity and duration, and exploring the relationship between these abilities and their vowel perception performance.

5. REFERENCES

[1] Best, C. (1995) ‘A direct realist view of cross-language speech perception’. In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in cross-language research.* York Press.

[2] Flege, J. (1995) ‘Second language speech learning: theory, findings, and problems.’ In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. York Press.

[3] Al-Tamimi, J (2007) Static and Dynamic cues in Vowel Production: a cross dialectal study in Jordanian and Moroccan Arabic

[4] Iverson, P. and Evans, G. (2009) ‘Learning English vowels with different first-language vowel systems II: auditory training for native Spanish and German speakers.’ *Journal of Acoustical Society of America*, 126 (2), pp 866-877.

[5] Van Hoof, J. and Verhoeven, J. (2011) ‘Intrinsic vowel F0, the size of vowel inventories and second language acquisition.’ *Journal of Phonetics*, 39, pp 168-177.

[6] Lee, W. (2011) ‘Dispersion and variability of vowels of different vowel inventory sizes.’ *Journal of Acoustical Society of America*. *Proceedings of Meetings on Acoustics,* 14, pp 1-9.

[7] Newman, D. and J. Verhoeven (2002) Frequency analysis of Arabic vowels in Connected Speech. *Antwerp Papers in Linguistics*. 100: 77-86.

[8] Ahmed, A. (2008) Production and Perception of Libyan Arabic Vowels. *PhD Thesis*. University of Newcastle.

[9] Tatham, M and Morton, K. (1988) The Perception and Production of Speech. A textbook.

[10] Peperkamp, S and Dupoux, E. (2002) ‘A Typological study of stress deafness’. *Lab Phonol* (7), pp 203–240.

[11] Fry, D. (1964) ‘The dependence of stress judgments on vowel formant structure’, In: S. Karger, Basel (eds.). *Proceedings of the 5th International Congress of Phonetic Science*. University of Munster, pp 306–311.

[12] Zhang, Y. and Francis, A. (2010) ‘The weighing of vowel quality in native and non-native listeners’ perception of English lexical stress’. *Journal of Phonetics*, 38, pp 260-271.

[13] Frost, D. (2011) ‘Stress and cues to relative prominence in English and French: a perceptual study.’ *Journal of International Phonetic Association*, 41, 1, pp 67-84.

[14] Chrabaszcz, A., Winn, M., Lin, C. & Idsardi, W. (2014) ‘Acoustic cues to the perception of word stress by English, Mandarin, and Russian speakers.’ *Journal of Speech, Language, and Hearing*. Vol. 57, Issue 4, pp 1468-1479.

[15]Cooper, N., Cutler, A.,& Wales, R. (2002). ‘Constrain ts of lexical stress on lexical access in English: Evidence from native and non-native listeners.’ *Language and Speech*, 45 (3), pp 207–228.

[16] Fry, D. (1955) ‘Duration and intensity as physical correlates of linguistic stress’. *Journal of Acoustical Society of America*, 27, 765.

[17] Fry, D. (1958) ‘Experiments in the perception of stress’. *Language and Speech*, 1(2), pp126–152.

[18] Lieberman, P. (1960) ‘Some acoustic correlates of word stress in American English’. *The Journal of Acoustical Society of America.* 32 (4), pp 451- 454.

[19] Morton, J. and Jassem, W. (1965) ‘Acoustic correlates of stress’. *Language and Speech*, (8), pp 148-158.

[20] Zuraiq, W. (2005) The Production of Lexical Stress by Native Speakers of Arabic and English and by Arab Learners of English. *A PhD Thesis*. University of Kansas.

[21] Bouchhioua, N. (2008) ‘The role of vowel quality in cuing stress and accent in Tunisian Arabic, native English, and L2 English.’ *In Speech Prosody*. *Fourth Conference on Speech Prosody, pp* 539-542.

[22] Bouchhioua, N. (2016a) ‘Typological variation in the phonetic realization of lexical and phrasal stress: Southern British English vs. Tunisian Arabic’.

[23] Bouchhioua, N. (2016b) ‘Stress and accent in Tunisian Arabic’. *Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics*.

[24] Zuraiq, W. and Sereno, J. (2007) ‘English lexical stress cues in native English and non-native Arabic speakers.’ *Ratio 1.*

[25] Almbark, R., Bouchhioua, N., & Hellmuth, S. (2014) ‘Acquiring the phonetics and phonology of English word stress: comparing learners from different L1 background.’ *Proceedings of the International Symposium of the Acquisition of Second Language Speech*. Concordia Working Papers in Applied Linguistics.

[26] Watson, JCE (2011) ‘Word stress in Arabic’. In: Ostendorp, M., Ewen, C., Hume, E. and Rice, K., (eds.) The Blackwell companion to phonology.

[27] Lin, Ch. (2018) The Perception and Production of Arabic lexical Stress by Learners of Arabic: A Usage-Based Account. *PhD dissertation*. University of Michigan.

[28] Othman, A. and Abubaker, S. (2021) ‘Investigating pronunciation difficulties encountered by Arab students in learning English sounds.’ *Global Libyan Journal* 50.

[29] Almbark, R. (2014) ‘Perception and production of SSBE vowels by foreign language learners: towards a foreign language model’. *Proceedings of the Int’l Symposium on the Acquisition of Second Language Speech Concordia Working Papers in Applied Linguistics* (5).

[30] Alshangiti, W. and Evans, B. (2018) ‘The perception and production of British English vowels and consonants by Arabic learners of English’. *Journal of Phonetics* (68), pp 15-31.

[31] Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4.

[32] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[33] Kuznetsova A, Brockhoff PB, Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26. doi:10.18637/jss.v082.i13.