

The perception and production of British English vowels and consonants by Arabic learners of English.

Bronwen G. Evans^{a*}, Wafaa Alshangiti^b,

*Corresponding author: bronwen.evans@ucl.ac.uk, Tel., +44 20 7679 4089

^a Dept Speech Hearing and Phonetic Sciences, Division of Psychology & Language Sciences, Chandler House, University College London, London, WC1N 1PF, United Kingdom.

^b King Abdulaziz University, English Language Institute, Jeddah, Saudi Arabia

Abstract

This study investigated the perception of British English vowels and consonants by native Saudi Arabic learners of English from a range of proficiency levels. Twenty-six participants completed consonant and vowel identification tasks in quiet and noise. To investigate if predicted difficulties with vowel perception were also present in production, participants also recorded vowels embedded in words and read a short story. The results demonstrated that all learners were better able to identify consonants than vowels in quiet and noise, with more experienced learners outperforming early learners. Although learners were likely able to rely on mapping non-native to native categories when identifying consonants, there was some evidence that they had started to establish new vowel targets. This appeared to start early in learning but even highly experienced learners continued to find vowels with no direct Arabic counterpart difficult. Additionally, there was some evidence for a link between perception and production: vowel perception was better in those who had more accurate production. Overall, the results shed light on problematic phonemic contrasts for Arabic learners, and suggest that though learners may be able to establish new phonetic categories early in learning, other contrasts continue to remain difficult even for highly experienced learners.

Keywords

Second language learning, Speech production, Speech perception

1.0 Introduction

It is well-known that early experience with a first language (L1) influences perception and production of a second-language (L2), such that the acquisition of non-native phonemes in adulthood can be impeded, particularly in cases where one or both target phonemes in the contrast are realised differently or do not occur in the learner's L1 (e.g., Best, 1994; Flege, 1995; Iverson et al., 2003). One possibility is that this is because experience with the L1 alters low-level processing, and that these changes interfere with adult learners' ability to alter existing representations and to form new categories for L2 sounds (see Iverson et al., 2003). For example, Japanese listeners are well-known to have difficulty acquiring the English /r/-/l/ contrast, likely because they are more sensitive to F2, a cue which is irrelevant for /r/-/l/ discrimination but which is associated with the Japanese flap /r /, than to F3 onset frequency, the cue used by native English speakers (see Iverson et al., 2003; Hattori & Iverson, 2009). Likewise, for vowels, L2 learners who use duration contrastively in their L1 are better at learning vowel contrasts that differ according to duration than those who do not use this cue in their L1 (e.g., Gottfried & Beddor, 1988; McAllister et al., 2002; Morrison, 2002; though see Bohn, 1995). This is consistent with the idea that language-specific perceptual processing modifies the relative salience of acoustic features, such that L2 features that are not contrastive in an individual's L1 are harder to perceive and produce.

Related to this, previous work has suggested that the relationship between the L1 and L2 phoneme systems affects learning. Flege's Speech Learning Model (SLM; Flege, 1995, 1999, 2002) hypothesizes that the L1 and L2 systems exist in the same phonological space, and that experience with an L1 plays a salient role in the way in which the L1 and L2 phonetic subsystems interact. In particular, L2 segments which are phonetically similar to L1 categories are thought to be assimilated into those L1 categories but those that are perceptually distinct from any L1 category are thought to be easier to learn, since they fall into relatively unoccupied regions in the phonological space (cf. Iverson & Evans, 2009). Similarly, Best's Perceptual Assimilation Model (Best, 1994, 1995) predicts that the difficulty in differentiating non-native phonemic contrasts is predictable from the basis of the relationship between the L1 and L2 phoneme inventories; for naïve perceivers, discrimination will be most difficult if both phonemes are assimilated equally well or poorly to the same L1 category, and best where two non-native phonemes are assimilated into two different L1 categories. For L2 learners in the process of

developing an L2 system, assimilation patterns are thought to be determined not just by L1-L2 relationships but also by how contrasting L2 phonemes relate to each other within the emerging L1-L2 phonological space (PAM-L2; Best & Tyler, 2007). This means that discrimination of contrasts in which one phoneme is perceived to be a good exemplar of a given L1 category (i.e., is perceptually assimilated) is predicted to be good and no new category will likely be formed. In cases where both L2 phonemes are perceived as equivalent to the same L1 phoneme but one is perceived to be a better fit than the other, discrimination should also be good, but with new category formation predicted only for the deviant phoneme. In contrast, where both L2 phonemes are assimilated to the same L1 category but are perceived to be equally good or equally poor instances of that category, discrimination of the L2 contrast will be poor initially and learning likely difficult. Lastly, where neither of the L2 phonemes is perceived as belonging to a particular L1 phoneme, i.e., the phonemes are *uncategorized*, then learning may be relatively easy depending on the relationships in the L1-L2 phonological system.

For languages such as Arabic with small vowel inventories (see e.g., Holes, 2004), where there are fewer opportunities to map L1 to L2 categories, it is possible that more L2 phonemes may fall into the uncategorized category (cf. Escudero & Williams, 2011), but that these phonemes may be uncategorized in different ways. Faris et al. (2016) examined the different ways in which native Egyptian Arabic (EA) listeners assimilated Australian English (AusE) vowels (19 vowels) to their L1 vowel categories (10 vowels), finding that there were indeed different observable patterns within uncategorized phonemes, and that these were related to the perceived relationships between the L1 and L2 phonetic and phonological categories. In their data, uncategorized vowels fell into 3 categories; (1) *focalized*, in which the L2 phoneme was perceived to be primarily similar to a single L1 category, but where responses were below their assumed threshold for categorization of 50% (e.g., AusE /i:/ which was primarily assimilated to EA /i:/); (2) *clustered*, in which the L2 phoneme was perceived to be similar to a small set of L1 categories (e.g., AusE /i/ which was split between EA /i:/ and /i/), and (3) *dispersed*, in which a range of L1 phonemes were used as responses because there was no good fit (e.g., AusE /ɜ:/). Based on their findings, the authors predict that for dispersed assimilations in which listeners are unable to detect L1 category invariant features, a new L2 category is likely to be formed because there will be no interference from previous L1 attunement. For focalized and clustered responses however, new category formation is predicted to depend on the degree of overlap with other L1

phonemes, such that a new category is only likely to be formed where the L2 focalized or clustered phoneme doesn't overlap with any other L2 category.

The present study aims to further investigate the acquisition of L2 phonemic contrasts and how this is affected by the relationship between the L1 and L2 by examining the perception and production of English vowels and consonants by Arabic learners of English living and working in London, UK. Although Arabic speakers potentially represent one of the largest groups of L2 English users and in many Arabic countries English is “viewed [...] as the language of technology, progress, and the future” (Nickerson and Camiciottoli, 2013; p.333), little previous research has investigated Arabic speakers' perception and production of English. What work there is has generally focussed on bilingual English-Arabic speakers. For example, Shafiro et al. (2012) tested early Arabic-English bilinguals (from different Arabic dialect backgrounds), and native English speakers of the English dialects spoken in the United Arab Emirates (UAE-English speakers), in their identification of American English vowels (/h/-V-/d/ words), and consonants (three vocalic contexts; /aCa/, /iCi/, /uCu/). Not surprisingly, participants performed very well. Consonant identification accuracy was 95% for Arabic-English bilinguals and 94% for the UAE-English speakers. Although all listeners found American English vowels that did not have a Modern Standard Arabic (MSA) counterpart (e.g., /ɑ/, /ɔ/, /æ/) more confusable than those that did, overall vowel identification was also high; 70% for the Arabic-English bilinguals and 80% for the UAE-English speakers. Given the much smaller vowel space of Arabic, it is perhaps somewhat surprising that participants performed so well in vowel identification. However, these participants were early bilinguals or native speakers of UAE-English dialects with high proficiency in English. Given previous work on L2 vowel acquisition and the recent findings reviewed above showing that AusE vowels were largely uncategorized by native Egyptian Arabic listeners¹, with the majority clustered or dispersed uncategorizations (Faris et al., 2016), it is thus highly likely that Arabic adult L2 learners of English (i.e., not early bilinguals) will have difficulties in accurately perceiving and producing English phonemes, in particular English vowels (cf., Best, 1994; Flege, 1995; Harnsberger, 2001; Iverson et al., 2003).

However, predicting exactly what difficulties an Arabic learner might have in terms of acquiring the English vowel and consonant phoneme inventory is not straightforward, as the

¹ Note that Faris et al (2016) do not specify whether or not their Arabic speakers were naïve or L2 learners of English, or give any details of their proficiency with English.

relationship between an L1 and L2 is somewhat more complicated in Arabic than in other languages. Arabic is a diglossic language, with a high and low variety. The high variety (Classical Arabic or MSA) is used in written forms and in formal settings, while the low variety (i.e., dialectal or colloquial Arabic) is used in daily conversations (Ferguson, 1959; Holes, 2004). Dialectal Arabic differs from classical Arabic in terms of its phonology, syntax, and lexicon. Recently the term Modern Standard Arabic (MSA) has emerged to refer to standard Arabic, a variety that uses the Standard Arabic lexicon, but preserves the phonological norms of the speaker's dialect (Watson, 2002). Consequently, there is a lot of variation between low varieties from different parts of the Arab-speaking world. Since the phonemic categories in different dialects will influence how listeners perceive or assimilate novel phonemes in relation to their native categories, and given the fact that other studies of Arabic phonetics have found that Arabic dialects typically differ from MSA (e.g., Bani-Yassin & Owens, 1987; Al-Tamimi, 2007), Arabic learners' difficulties will thus likely vary according to their dialect background.

Our study investigates the perception of English vowels and consonants by native Saudi Arabic learners of English with a range of abilities in English (Experiment 1). In a small pilot study, we explored the Arabic vowel and consonant variants used in speakers from the 2 largest cities in Saudi Arabia, Riyadh and Jeddah. Twelve Saudi Arabic speakers (6 from Jeddah, 5 male) aged 19-35 years (median 27 years old) were recorded completing various different tasks that elicited Arabic in different speech styles; reciting the Qur'an, reading and retelling a story, naming pictures in their dialect, and completing a sociolinguistic interview. In addition to using the standard 28 MSA consonants (e.g., Holes, 2004), these Saudi Arabic speakers also used other variants. They used the low variant /g/ in informal settings for the high variant /q/, and /dʒ/ in formal speech and when reciting the Qur'an, while in less formal settings they used the low variant /ʒ/ in place of /dʒ/. We expect that this will facilitate perception of English consonants which map well to the Saudi Arabic consonant inventory. In terms of vowels, Standard Southern British English (SSBE) is typically described as having 12 monophthongs, and 8 diphthongs (e.g., Wells, 1982), while MSA is described as having 3 tense-lax monophthong pairs [/i/-i:/, /a/-a:/, /u/-u:/; e.g., Holes (2004)]. Our speakers used a similar vowel inventory to that of MSA in less formal settings, though they tended to have more centralized productions overall. Despite these small differences, we predict that their small vowel inventory will make it hard to map one

English vowel to one Arabic vowel, and that overall, perception of vowels will therefore be less accurate.

Even so, Saudi Arabic learners are frequently exposed to English from a young age in their home country, through both education and more informally through the media, and so one possibility is that our participants, even those considered to have little direct experience with native English speakers (e.g., by living in the UK) might perform well in phoneme identification tasks. Consequently, to avoid the possibility of ceiling effects, participants completed natural vowel and consonant phoneme identification tasks in quiet and in noise. Given that the Arabic vowel inventory uses duration contrastively, we also investigated whether participants relied more on this cue when identifying English vowels by including duration-equated vowels in the vowel identification in noise task. For comparison, native English controls completed all vowel and consonant identification in noise tasks. Additionally, to investigate how predicted difficulties with vowel perception might affect vowel production, the same Arabic participants were recorded producing the /h/-V-/d/ vowel stimuli they had identified in the vowel identification task and a short passage (Experiment 2). English native speakers then identified the vowels and rated their speech for accentedness.

2.0 Experiment 1: Perception of English Vowels & Consonants

2.1 Method

2.1.1 Participants

A total of 35 participants took part in the study. Twenty-six Saudi Arabic (SA) speakers, born and raised in Jeddah (N=14) and Riyadh (N=12), were tested in their perception of English vowels and consonants in quiet and noise and production of English vowels (Experiment 2). Nine native Standard Southern British English (SSBE) listeners were recruited as controls and completed a subset of the perception tasks to give normative data. These SSBE listeners also completed identification and ratings tasks for Arabic participants' English production (Experiment 2). All participants were 18-35 years old (median 26 years), reported no speech or hearing problems and were resident in London at the time of testing. All participants volunteered to take part in the study and were compensated for their time.

SA speakers were recruited to cover a range of proficiency levels and had acquired English at different ages. Participants began learning English when they were 2-23 years old (median 11 years), and had 3 months-9 years' experience of living in the UK (median 3 years). They had not spent any significant amount of time in any other English-speaking country. All participants had learned English primarily through school and university education, where it is compulsory for SA students to learn English, but were not studying English formally at the time of testing. In order to obtain a general measure of participants' overall English language proficiency independent of their perceptual abilities, participants completed Part 1 of the *Oxford Placement Test 1* (Allan, 1992). This sub-section comprises a written test which provides a measure of learners' comprehension and grammatical skills. Scores ranged from 34% - 94%, with an average of 60%, indicating that students ranged in their ability from a lower intermediate level of English competence (i.e., a functional, but not fluent command of English) to Advanced level (i.e., a fluent command of English). Although there was a strong correlation between performance on this task and Age of Acquisition of English ($r = -0.69, p < 0.001$), with those who began learning English earlier achieving higher scores (cf. Flege, 1999) there was no relationship between performance and Length of Residence in the UK ($r = 0.28, p > 0.05$; cf. Jia et al., 2006)

2.1.2 Stimuli and Apparatus

Consonant perception (Quiet and Noise). A male monolingual SSBE speaker recorded the English consonants in two VCV contexts. The speaker was aged 40 years, had been born and raised in the south of England and was living in London at the time of the recording. The speaker recorded three versions of each consonant / p b m f v θ ð t d n s z ʃ ʒ ʧ ʤ g k ŋ l w r h/ in two vocalic contexts /iCi/ and /aCa/, with stress on the second syllable, embedded in the carrier sentence "Say __ again" giving a total of 138 tokens. The vocalic contexts were varied because this has been shown to affect phonemic perception in L2 learners (cf. Strange et al., 2007). To avoid list intonation, all tokens were displayed one at a time on a computer screen in a random order during the recording session. Recordings were made in a sound-attenuated audio booth using a Røde NT-1A microphone connected to an Edirol UA-25 sound card, at a sampling rate of 44.1 kHz, 16-bit resolution. Each word was checked for clarity and the best production was selected, down-sampled to 22,050Hz and amplitude-normalized to 70 dB.

Stimulus sets for identification in quiet used these selected recordings unaltered. The stimulus sets for identification in noise were created by mixing the selected recordings with speech-shaped noise (S. Rosen, UCL) generated by a Wandel and Goltermann RG-1 noise generator at three signal-to-noise (SNR) ratios (0, -5, and -10 dB). In order to create the different noise conditions, the root mean square (RMS) amplitude of the stimulus and noise were determined and scaled to fit the SNR condition. They were then combined through addition at the three SNRs using an automated script in Praat (Boersma & Weenink, 2013). Finally, all stimuli files were equalized for intensity at 70 dB.

Vowel perception (Quiet and Noise). The same male SSBE speaker recorded the vowel stimuli. Three versions of 17 vowels covering the majority of the SSBE vowel space were recorded; /i: ɪ e æ a: ɒ ɔ: u: ʊ ʌ ɜ: eɪ aɪ aʊ əʊ eə ɔɪ/. Vowels were produced in the carrier sentence "Say _____ again" in a /h/-V-/d/ context, giving the words; *heed, hid, head, had, hard, hod, hoard, who'd, hood, hud, heard, hayed, hide, how'd, hoed, haired, hoyed*. Recordings were made under identical conditions and using the same equipment as for the consonant recordings. Again, each word was manually checked for clarity and the best production was chosen, down-sampled to 22050Hz and amplitude-normalized to 70 dB.

The selected stimuli were used to create stimuli for three experimental conditions: quiet, natural vowels in noise, and duration-equated vowels in noise. The latter condition was included to test the use of duration as a cue in vowel identification. The Arabic vowel inventory includes short-long pairs, and so it is possible that Arabic learners are able to make use of duration as an L1 cue when identifying English vowels (cf. Gottfried & Beddor, 1988; MacAllister et al., 2002; Morrison, 2002). Duration equated vowels were created using PSOLA implemented in Praat (Boersma & Weenink, 2013). The duration of the /h/ closure, the vowel, and the /d/ closure were averaged across all vowels for the talker, and these values were used for all words. The duration of the vowel portion was set to the average value calculated across all vowels, 302 ms (see Appendix B for natural vowel durations). Individual words were excised from the carrier sentence and processed individually before being spliced back into a single carrier sentence. To create the stimuli for the different noise conditions, recordings were equated for amplitude and then speech-shaped noise was added to the natural and duration-equated recordings to create three SNRs (0, -5, -10 dB), in the same way as for the consonants. Finally, all stimuli files were

equalized for intensity at 70 dB. Stimulus sets for identification in quiet used the selected recordings unaltered.

2.1.3 Procedure

All perception experiments were carried out in sound-attenuated audio-booths at UCL Language Sciences, Chandler House. Stimuli were presented over Sennheiser HD 555 headphones and both stimuli presentation and response collection were controlled using Praat (Boersma & Weenink, 2013). SA participants completed Part 1 of the Oxford Proficiency Test using paper and pen at the end of the testing session.

Consonant identification in quiet. This task was completed by SA participants only. Participants heard natural recordings of the English speaker for the consonants in the two vocalic contexts, /aCa/ and /iCi/, in the carrier sentence *Say__ again* (e.g., “*Say /aka/ again*”, “*Say /aɟʒa/ again*”). They were asked to give a closed-set identification response with all 23 words as response options. To give their response, participants mouse-clicked on a button which listed the consonant and a common English word that began with that consonant, e.g., “*B as in **B**ear*”, “*SH as in **S**harp*” (see Appendix A for a list of keywords). Words were selected to be high frequency and pilot testing confirmed that they were likely to be familiar to all participants regardless of L2 proficiency. Before completing the experiment, participants were familiarized with the task and materials, and in particular with words where the acoustic-orthographic correspondence is not transparent, (e.g., ‘th’ can be produced as /ð/ as in *faTHer*, or as /θ/ as in *THeatre*).

Participants identified three repetitions of each consonant in each context, giving a total of 138 responses (23 consonants x 3 repetitions x 2 vocalic contexts; 69 stimuli for each vocalic context). Stimuli were blocked by context and the order of presentation within each block randomized. To control for any training or order effects, the order of presentation (i.e., /iCi/-/aCa/ or /aCa/ - /iCi/) was counterbalanced across participants. The test was self-paced with a break mid-way through (i.e., after 69 stimuli). Participants heard each stimulus once and received no feedback.

Consonant identification in noise. This task was completed by SA and SSBE participants. Participants identified two repetitions of each consonant in the two vocalic contexts (/aCa/ and /iCi/) at three different SNRs: 0 dB, -5dB and -10dB. This gave a total of 46 stimuli for each

vocalic context, and a total of 92 stimuli per noise condition. The experiment was blocked by noise level and the order of presentation of the blocks randomized to control for any learning effects. Additionally, the order of presentation of the stimuli was randomized within each block. The test was self-paced with a break mid-way through (i.e., after 138 stimuli). The experiment was the same as the consonant identification in quiet in all other respects.

Natural vowel identification in quiet. Only SA participants completed this condition. Participants listened to recordings of the vowels in /h/-V-/d/ words in the carrier sentence “Say ___ again”, and gave a closed-set identification response from the 17 test words. To give their response, participants mouse-clicked on a button which listed the /h/-V-/d/ word and a with common English rhyme word, (e.g., *heed* as in *seed*, *hud* as in *cut*; see Appendix B for a list of keywords and rhyme words). As for the consonantal stimuli, pilot testing confirmed that these were likely to be familiar to all participants regardless of L2 proficiency. Participants identified three repetitions of each vowel in a randomized order, giving a total of 51 trials. The test was self-paced with no break. Participants heard each word once and received no feedback.

Natural and duration-equated vowel identification in noise. This task was completed by SA and SSBE participants. Participants identified natural and duration-equated vowels separately. For each set, they identified two repetitions of each vowel at the three different SNRs (0 dB, -5 dB and -10 dB), giving a total of 102 stimuli (17 vowels x 3 SNR levels x 2 repetitions, giving 34 stimuli per noise level). The experiment was blocked by noise level and vowel type; listeners identified natural then duration-equated vowels. The order of presentation of the stimuli was randomized within each block. Responses were collected using the same procedure used in vowel identification in quiet. The test was self-paced with a break mid-way through (i.e., after 51 stimuli). The experiment was the same as the natural vowel identification in quiet in all other respects.

2.2. Results

Results were analyzed for each task separately. To facilitate comparison of groups based on proficiency with English, we performed a median split on the data based on participants’ score on the Oxford Placement Test 1 (Allan, 1992; see Section 2.1.1). Participants were assigned to the HP group (N=13, Mean score = 35.9/50, SD = 5.35) if their score was higher than

or equal to the median score of 29.5, and the LP group (N=13, Mean score = 23.08/50, SD = 3.64) if their score was lower than the median. The HP group thus all had an advanced level of English competence, whilst the LP group had a lower intermediate level of English competence.

2.2.1. Consonants

Consonant identification in quiet. As displayed in Fig. 1, overall average accuracy was high in both groups (HP; 86.5%, LP; 73.5%) and similar across vocalic contexts (HP; 86% for /aCa/ and 87% for /iCi/, LP; 73% for /aCa/ and 74% for /iCi/), though as expected, the LP group appeared to perform more poorly overall than the HP group. As performance did not differ across vocalic contexts, an average score per participant was calculated and this was used in all subsequent analyses.

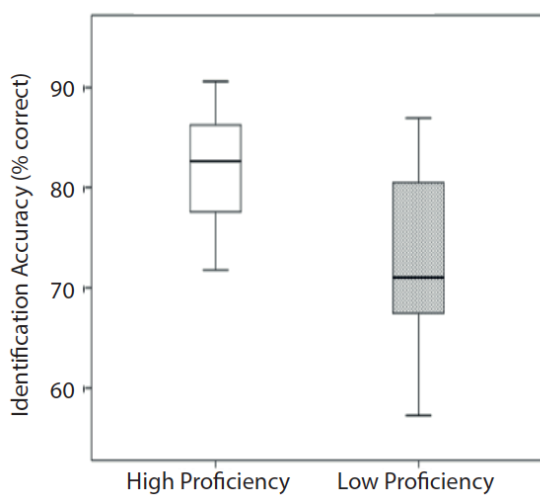


Figure 1. Boxplot showing the consonant identification accuracy (percentage correct) in quiet averaged across vocalic contexts and split into high proficiency (left box) and low proficiency (right box) groups. The lower and upper whiskers represent the first and last quartiles respectively.

An independent samples t-test confirmed that proficiency level was a significant factor in L2 listeners' ability to categorize L2 phonemes, [$t(24) = 3.6, p < 0.05, df = 24$]: HP listeners were more accurate than LP listeners. A series of analyses further investigated the perceptual confusion patterns and whether or not these differed according to proficiency. Table 1 displays the confusion matrix for the LP group averaged over both vocalic contexts. Participants were

very accurate with many phonemes but performed more poorly with the affricates /tʃ/ (33%) and /dʒ/ (31%), the fricative /ʒ/ (28%), the dental fricatives /θ/ (54%) and /ð/ (64%), and the velar nasal /ŋ/ (47%). Identification of /p/ was good (74%), even though this is not a native phoneme, but /b/ was less well identified (68%) and was most frequently confused with /p/, indicating that learners struggled with the voicing contrast. The pattern of results for the HP group was similar (see Table 2); they performed very accurately with most phonemes but also more poorly with the affricates /tʃ/ (76%) and /dʒ/ (59%), the fricative /ʒ/ (33%), and the velar nasal /ŋ/ (60%). Performance on the dental fricatives /θ, ð/ was also slightly lower than for other well-identified phonemes (74% and 79% respectively) but was still relatively high.

Table 1: Consonant Confusion matrix for the low proficiency (LP) group; the stimuli are in rows, and the responses (in percent) in columns. Empty cells show that there were no responses for this stimulus-response combination. As there were no significant differences in performance in the difference vowel contexts, responses are averaged over both vocalic contexts.

| Stimulus | Response | | | | | | | | | | | | | | | | | | | | | | | |
|----------|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|-----|----|----|
| | p | b | m | f | v | θ | ð | t | d | n | s | z | ʃ | ʒ | tʃ | dʒ | k | g | ŋ | h | l | ɹ | w | |
| p | 74 | 17 | | | 4 | | 1 | | | | | | | 1 | | | | | | | | | 3 | |
| b | 27 | 68 | | | 4 | | | | 1 | | | | | | | | | | | | | | | |
| m | | | 79 | | | 1 | | 1 | | 11 | | 1 | | | | 4 | 1 | 1 | | 1 | | | | |
| f | | | | 82 | 8 | 4 | 5 | | 1 | | | | | | | | | | | | | | | |
| v | | 1 | | 8 | 87 | | 3 | | | | | 1 | | | | | | | | | | | | |
| θ | | | | 27 | 1 | 54 | 14 | 1 | | | | 3 | | | | | | | | | | | | |
| ð | | | | | 9 | 19 | 64 | | | 1 | | 1 | 1 | 1 | | 4 | | | | | | | | |
| t | | | | | | | 8 | 92 | | | | | | | | | | | | | | | | |
| d | | | | | | | 1 | 1 | 87 | | | | | | | | | 8 | 3 | | | | | |
| n | | | | | | | | | | 79 | | | | | | | | | 21 | | | | | |
| s | | | | | | | | | | | 71 | 1 | 3 | | 24 | | | | 1 | | | | | |
| z | | | | | 1 | | | | | | 3 | 93 | | | | | | | | | | | | |
| ʃ | | | | | | | | | | | | | 74 | 1 | 23 | 1 | | | | 1 | | | | |
| ʒ | | | | | | | | | | | | | | 28 | | 21 | 1 | 42 | 8 | | | | | |
| tʃ | | | | | | | | | | | | | 29 | 5 | 33 | 10 | | 17 | 5 | | | | 1 | |
| dʒ | | 1 | | | | | | | | | | | | 23 | | 31 | | 36 | 9 | | | | | |
| k | | | | | 1 | | 1 | | | | | | | | 4 | | 91 | 3 | | | | | | |
| g | | | | | 3 | | | | 3 | | | | | | | | 1 | 86 | 7 | | | | | |
| ŋ | | | 3 | | | | | | 1 | 47 | | | | | | 2 | | | 47 | | | | | |
| h | | | | 2 | 1 | | | 3 | | | 1 | | 4 | | | 1 | | | | 87 | | | 1 | |
| l | | | | | | | | | | | | | | | | | | | | | | 100 | | |
| ɹ | | | | | | | | | | | | | | | | | | | | | | | 99 | 1 |
| w | 1 | | | | | | | | | | | | | | | 1 | | | | | | | 13 | 85 |

Table 2: Consonant Confusion matrix for the high proficiency group (HP); the stimuli are in rows, and the responses (in percent) in columns. Empty cells show that there were no responses for this stimulus-response combination. As there were no significant differences in performance in the difference vowel contexts, responses are averaged over both vocalic contexts

| Stimulus | Response | | | | | | | | | | | | | | | | | | | | | | |
|----------|----------|----|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|----|----|---|----|------|
| | p | b | m | f | v | θ | ð | t | d | n | s | z | ʃ | ʒ | tʃ | dʒ | k | g | ŋ | h | l | ɹ | w |
| p | 86 | 8 | | | | | | | | | | 1 | 1 | 3 | | | 1 | | | | | | |
| b | 7 | 91 | | | | | | | | | | | 1 | | | 1 | | | | | | | |
| m | | | 100 | | | | | | | | | | | | | | | | | | | | |
| f | | | | 90 | 2 | 3 | 5 | | | | | | | | | | | | | | | | |
| v | | 1 | | | 91 | | 7 | | | | | | 1 | | | | | | | | | | |
| θ | | | | 20 | | 74 | 6 | | | | | | | | | | | | | | | | |
| ð | | | | | 9 | 9 | 79 | | | | | | | | | | | 2 | | | | | 1 |
| t | | | | | | 4 | | 95 | | | | | | | | | | | | | | | 1 |
| d | | | | | | | | | 92 | | | | | | | | | 8 | | | | | |
| n | | | | | | | | | | 94 | | | | | | | | | 6 | | | | |
| s | | | | | | | | | | | 90 | | | | 7 | | | | | | | | 3 |
| z | | | | | | | | | | | 1 | 98 | | 1 | | | | | | | | | |
| ʃ | | | | | | | 1 | | | | 1 | | 87 | 1 | 10 | | | | | | | | |
| ʒ | | | | | | | | | | | | | | 33 | | 45 | | 22 | | | | | |
| tʃ | | | | | | | | | | | | | 13 | 5 | 76 | 1 | | 3 | | 2 | | | |
| dʒ | | | | | | | | | | | | | | 17 | 1 | 59 | | 23 | | | | | |
| k | | | | | | | | | 1 | | | | 1 | | | | 98 | | | | | | |
| g | | | | | | | | | | | | | | | | | | 100 | | | | | |
| ŋ | | | | | | | | | | 35 | | | | | | | 1 | 4 | 60 | | | | |
| h | | | | | | | | | | | | | | | | | | | | 99 | | | |
| l | | | | 1 | | | | | | | | | | | | | | | | | | 99 | |
| ɹ | | | | | | | | | 2 | | 1 | 1 | | | | | | | | | | | 96 |
| w | | | | | | | | | | | | | | | | | | | | | | | 1 99 |

Separate hierarchical cluster analyses for the HP and LP groups were used to analyze the confusion patterns. For the LP group, there were five distinct clusters, the first containing the affricates, postalveolar fricatives and closest voiced stop, /g/, the second made up of the alveolar and velar nasals, the third containing the dental and labio-dental fricatives, and lastly, a cluster made up of the remaining phonemes. Within each of these clusters, certain groups of consonants were highly confusable; /dʒ/ and /ʒ/ were the most confusable and joined to form the first cluster. The alveolar nasal /n/ and the velar nasal /ŋ/ formed the second cluster. Within the third cluster, the dental fricatives, /θ/ and /ð/ were highly confusable. In the final cluster, /tʃ/ and /ʃ/ were highly confusable, as were the bilabial plosives /b/ and /p/. The cluster diagram for the HP group displayed some differences. There were three clusters, one containing /dʒ/, /ʒ/ and /g/, the second containing the nasals /n/ and /ŋ/, and the third the remaining consonants. As for the LP participants, the analysis showed that the voiced affricate /dʒ/ and the voiced fricative /ʒ/ were

the most confusable phonemes and joined to form the first cluster, followed by the alveolar nasal /n/ and the velar nasal /ŋ/.

Consonant identification in noise. Fig. 2 displays the English consonant identification accuracy in noise for each group (native SSBE, and HP and LP Saudi Arabic listeners) averaged across the two vocalic contexts. As expected, all listeners performed more poorly at higher noise levels, with all listener groups equally affected by noise. Performance appeared to be affected by proficiency; SSBE listeners performed best, followed by HP and then LP listeners. Interestingly, vowel context did not appear to affect performance except in the high noise condition (-10dB) where SSBE listeners showed a much larger drop in performance for the /iCi/ than for the /aCa/ context.

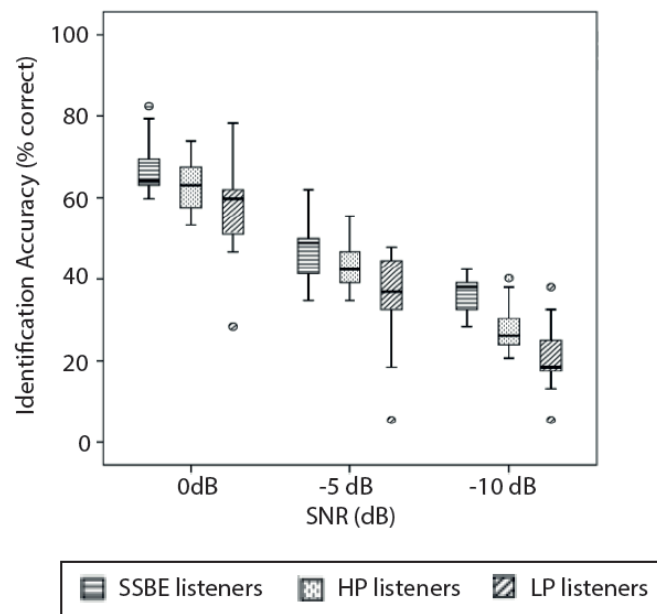


Figure 2: Boxplot to show consonant identification (percentage correct) in three different noise levels (0, -5, -10 dB) for three groups, native listeners (SSBE), and high and low proficiency (Saudi Arabic) listeners, averaged across vocalic contexts. The lower and upper whiskers represent the first and last quartiles respectively, with outliers represented by shaded circles.

These observations were tested using a repeated measures ANOVA with noise level (0dB, -5dB, -10dB) and vowel context (/aCa/, /iCi/) coded as within-subjects factors, and group (SSBE, HP, LP) as a between-subjects factor. As expected, the main effect of noise was significant [$F(2,64)=287.13, p<.001$], confirming that overall performance decreased as the noise

level increased (0dB; 61%, -5dB; 41%, -10dB; 28%). There was also a significant main effect of group, $[F(2,32)=7.66, p<0.01]$; average overall performance accuracy across noise conditions for the SSBE listeners was higher (49.9%) than for the HP (44.8%) and LP group (36.7%). A series of post-hoc independent samples t-tests revealed that this effect was driven largely by lower performance in LP listeners in comparison to SSBE and HP listeners. SSBE and HP listeners performed similarly in all conditions, $p >.05$, except for the -10 dB /aCa/ condition, in which SSBE listeners outperformed HP listeners $[t(20) = 5.98, p<.001]$. There was no difference in performance between SSBE and LP listeners in the 0dB and -5dB /aCa/ conditions, $p >.05$, but SSBE listeners were significantly better than LP listeners in all other conditions [/aCa/ at -10dB, $t(20)=4.71, p<.001$, /iCi/ at 0dB, $t(20)=3.18, p<.01$, /iCi/ at -5dB, $t(20) = 3.26, p<.01$, /iCi/ at -10dB, $t(20)=2.68, p<.05]$. There was no interaction between group and noise, $p>.05$, confirming that although LP listeners performed more poorly than native and HP listeners, all were similarly affected by noise (see also van Dommelen & Hazan, 2010).

Additionally, there was a significant effect of vowel context, $[F(1,32)=13.53, p<.001]$ and a three-way interaction of vowel context, noise and group $[F(4,64)=4.19, p<.01]$. Further exploration of the data indicated that there was a tendency for all participants to perform better overall with the /aCa/ context, but a series of post-hoc paired samples t-tests comparing each listener group's performance on each vowel context at each noise level (Bonferroni corrected for multiple comparisons) revealed that this was limited to SSBE listeners in the highest noise condition. Both HP & LP listeners performed similarly with the /iCi/ and /aCa/ contexts at each noise level, and SSBE listeners showed no difference in performance at 0dB and -5 dB SPL, $p>.016$. However, at -10 dB SPL, SSBE listeners performed better with the /aCa/ context, $t(8) = 6.54, p<.001$.

2.2.2 Vowels

Vowel identification in quiet. Fig.3 displays the accuracy for English vowel identification in quiet for HP and LP listeners. As expected HP listeners performed better than LP listeners. An independent samples t-test revealed that there was a significant difference between the HP and LP group $[t=2.72, p<.05, df=24]$, confirming that HP listeners identified English vowels more accurately than the LP group (average ID, HP: 69%; LP: 47%).

A series of analyses investigated whether patterns of perceptual confusions also differed according to proficiency. Table 3 displays the confusion matrix for the LP group. Participants were highly accurate with some phonemes, e.g., /i:/ (*heed*; 74%), /æ/ (*had*; 79%) /ɑ:/ (*hard*; 85%), but performed particularly poorly with the following vowels; the front-mid vowel /ɪ/ (*hid*; 8%), the open-back vowel /ɒ/ (*hod*; 3%), the mid closing diphthong /əʊ/ (*hoed*), and the central diphthongs /eə/ (*haired*; 18%) and /ɜ:/ (*hoed*; 18%). They also had marked difficulties with the high-back vowel /u:/ (*who'd*; 36%), and the low central vowel /ʌ/ (*hud*; 31%). Table 4 displays the confusion matrix for the HP group. These participants had fewer difficulties overall, but still found some of the same vowels problematic; performance was relatively poor for the front-mid vowel /ɪ/ (*hid*; 44%), the open-back vowel /ɒ/ (*hod*; 31%) and the low central vowel (/ʌ/ *hud*: 44%), and the central diphthong /eə/ (*haired*; 31%).

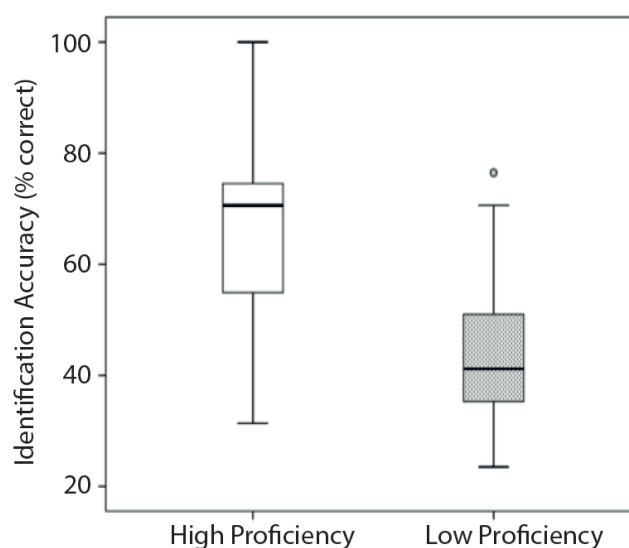


Figure 3: Boxplot to show the vowel identification accuracy (percentage correct) for HP and LP groups. High proficiency learners performed better overall than did low proficiency learners. The lower and upper whiskers represent the first and last quartiles respectively, with outliers represented by shaded circles.

Table 3: Vowel confusion matrix for the LP group listeners. The stimuli are in rows, and the responses (in percent) in columns. Empty cells show that there were no responses for this stimulus-response combination.

| Stimulus | Response | | | | | | | | | | | | | | | | |
|-------------|----------|-----|------|-----|------|-----|-------|-------|------|-----|-------|-------|------|-------|------|--------|-------|
| | heed | hid | head | had | hard | hod | hoard | who'd | hood | hud | heard | hayed | hide | how'd | hoed | haired | hoyed |
| heed /i:/ | 74 | 8 | 3 | | | | | | | | | 3 | 10 | | 2 | | |
| hid /ɪ/ | 3 | 8 | 72 | 5 | | | | | | 2 | 3 | 5 | | | | | 2 |
| head /e/ | 5 | 10 | 69 | 10 | | | | | | | | 3 | 3 | | | | |
| had /æ/ | | | | 79 | 2 | | | | | 13 | | | 3 | 3 | | | |
| hard /ɑ:/ | | | | 4 | 85 | 3 | 3 | | | 3 | 2 | | | | | | |
| hod /ɒ/ | | 2 | | 38 | 3 | 5 | | | 5 | 18 | 3 | | | 3 | 13 | 10 | |
| hoard /ɔ:/ | | | | 2 | 5 | 62 | 2 | 5 | 3 | 3 | | | | 10 | | | 8 |
| who'd /u:/ | | | | | 5 | 36 | 54 | | | | | | | 5 | | | |
| hood /ʊ/ | | | | | | 10 | | 23 | 51 | 8 | | | | | 8 | | |
| hud /ʌ/ | | 2 | 5 | 21 | 15 | | | | | 31 | 21 | 2 | | | 3 | | |
| heard /ɜ:/ | | | 5 | | 23 | | 3 | | | | 44 | 2 | | | | 23 | |
| hayed /eɪ/ | | | 18 | | 8 | | | | | | | 59 | 4 | | | 3 | 8 |
| hide /aɪ/ | | 26 | 3 | | | | | | | | | 18 | 46 | | | 2 | 5 |
| how'd /aʊ/ | | | | | 8 | 3 | 10 | 13 | | | | | 3 | 59 | 4 | | |
| hoed /əʊ/ | | | | | 8 | | 4 | 36 | 8 | | | | | 26 | 18 | | |
| haired /eə/ | | 8 | 20 | 20 | 3 | | | | | | 28 | 3 | | | | 18 | |
| hoyed /ɔɪ/ | | | | | | | 3 | | | | 2 | 5 | | 5 | 26 | 3 | 56 |

Separate hierarchical cluster analyses for the HP and LP groups were used to analyse the confusion patterns. For the LP group there were 4 distinct clusters; (1) the close-mid contrast *hid-head*, (2) the remaining high front vowels, including the front closing diphthongs /eɪ/, /aɪ/ and /ɔɪ/, (3) the high-back vowels, including the high-back closing diphthongs /əʊ/ and /aʊ/, and (4) the central and low back vowels. Within these clusters, certain pairs of vowels were highly confusable; /ɪ/-/e/ was the most confusable contrast and joined to form the first cluster, followed by the high back vowel cluster which was made up of 2 highly confusable pairs, /u:/-/ʊ/ and /əʊ/-/aʊ/. Within the central and low back vowel cluster, the /ɒ/-/ʌ/ and /ɜ:/-/eə/ contrasts were also highly confusable.

The analysis for the HP group showed a two-way split between front and back vowels with smaller clusters within these larger groups. The /ɒ/-/əʊ/ contrast was the most confusable and joined to form the first cluster, followed by the /ɪ/-/e/ contrast and the /ɜ:/-/eə/ contrast. This was followed by the /u:/-/ʊ/ and /ɔ:/-/aʊ/ contrasts within the back vowel cluster. Although the

HP group had fewer difficulties overall, they shared some of the same vowel confusions with the LP group; /ɪ/-/e/, /ɜ:/-/eə/, /u:/-/ʊ/ and /ɒ/-/ʌ/-/əʊ/.

Table 4: Vowel confusion matrix for the HP group listeners. The stimuli are in rows, and the responses (in percent) in columns.

| Stimulus | Response | | | | | | | | | | | | | | | | |
|-------------|----------|-----|------|-----|------|-----|-------|-------|------|-----|-------|-------|------|-------|------|--------|-------|
| | heed | hid | head | had | hard | hod | hoard | who'd | hood | hud | heard | hayed | hide | how'd | hoed | haired | hoyed |
| heed /i:/ | 85 | | 8 | | | | | | | | 4 | 3 | | | | | |
| hid /ɪ/ | | 46 | 44 | 2 | | 3 | | | | | | | 5 | | | | |
| head /e/ | | 10 | 85 | 3 | | | | | | | | | 2 | | | | |
| had /æ/ | | | | 87 | | 5 | | | | 8 | | | | | | | |
| hard /ɑ:/ | | | | | 85 | 3 | | | 2 | | 5 | | | | | | 5 |
| hod /ɒ/ | | | | | 8 | 33 | 5 | | 15 | 10 | | | | 8 | 21 | | |
| hoard /ɔ:/ | | | 2 | | | | 67 | 3 | 3 | | | | | 15 | 5 | | 5 |
| who'd /u:/ | | | | | | 2 | 2 | 59 | 21 | | | | | 8 | 3 | | 5 |
| hood /ʊ/ | | 3 | | 3 | | 15 | 3 | 5 | 69 | | | | | | | 2 | |
| hud /ʌ/ | | 3 | | 8 | 3 | 13 | | | 8 | 49 | 10 | 2 | | 2 | 2 | | |
| heard /ɜ:/ | | | 3 | | 15 | | 3 | | | | 79 | | | | | | |
| hayed /eɪ/ | | | 3 | | | | | | | | 5 | 87 | 3 | | | | 2 |
| hide /aɪ/ | | | 8 | | | | | | | | | 2 | 90 | | | | |
| how'd /aʊ/ | | | | | | 3 | 2 | 2 | 3 | | | | | 79 | 8 | | 3 |
| hoed /əʊ/ | | | | | | | | 8 | 18 | | | 3 | | 18 | 51 | | 2 |
| haired /eə/ | | | 21 | 8 | 3 | | 2 | | | | 33 | 3 | 2 | | | 28 | |
| hoyed /ɔɪ/ | | | | | | | 3 | | | | | 5 | | 10 | | | 82 |

Vowel identification in noise; natural vs. duration equated vowels. Fig. 4 displays the accuracy performance for the three listener groups (native SSBE and HP and LP Saudi Arabic listeners) for natural and duration equated vowels. As expected, non-native listeners appeared to perform more poorly than SSBE listeners, with LP listeners performing more poorly than HP listeners. Since duration is contrastive in Arabic, we had hypothesized that Arabic listeners, especially the LP group, might rely on duration more when identifying vowels that are not present in their L1 (e.g., *hid*) than would SSBE listeners, who are thought to rely more on spectral rather than duration information (see e.g., Escudero & Boersma, 2004). However, inspection of the data revealed that both SSBE and HP listeners had a relatively large drop in performance from natural to duration-equated vowels (SSBE listeners: 75% for natural vowels, 66% for duration equated vowels; HP listeners: 55% for natural vowels, 42% for duration

equated) compared to LP listeners who performed equally poorly in both conditions (natural vowels 31%, duration-equated 28%).

These observations were tested in a repeated measures ANOVA with duration (natural, duration-equated), and noise (0dB, -5dB, -10dB) as within-subjects-factors, and group (SSBE, HP, LP) as a between-subjects factor. As expected, there was a main effect of noise [$F(2,64)=21.7, p<.001$], confirming that for all participants, performance dropped as the noise level increased (0dB, 59%; -5dB, 5%; -10dB, 35%). Additionally, there was a main effect of group [$F(1,32)=31.78, p<.001$]; overall performance for the SSBE listeners was higher (71%) than for the HP (46%) and LP groups (29%).

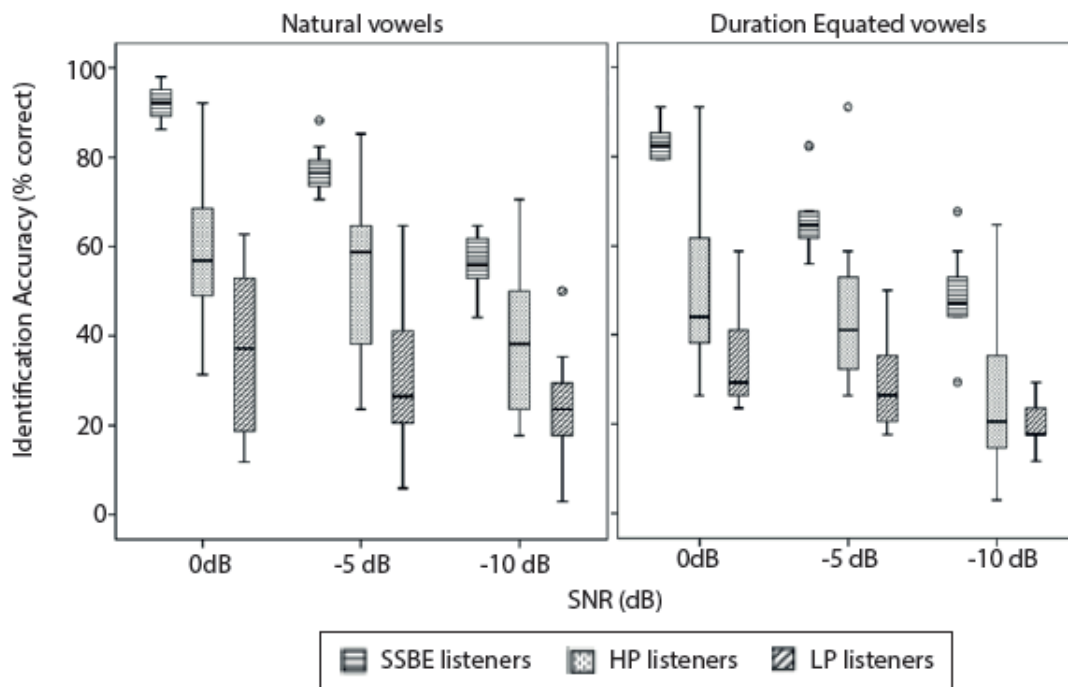


Figure 4: Boxplots showing the overall vowel identification scores (percentage correct) for the three groups (N, HP, and LP) in natural vowels, and in the duration equated condition at the three noise levels (0, -5, and -10 dB). The lower and upper whiskers represent the first and last quartiles respectively, with outliers represented by shaded circles.

There was a significant interaction between noise and group [$F(4,64)=13.62, p<.001$]. Inspection of the data revealed that this was because SSBE listeners were more affected by the

higher noise levels than were the non-natives, who performed more poorly at the easier noise level. There was a main effect of duration [$F(1,32)=17.51, p<.001$]; overall identification of natural vowels averaged across all listeners was higher for natural vowels (52%) than for the duration equated vowels (44%), indicating that participants found the natural vowels easier to identify. However, there was no statistically significant interaction of duration and listener group ($p > .05$), indicating that there was no reliable evidence that different listener groups behaved differently with natural and duration-equated vowels in noise.

2.3 Discussion

The results demonstrated that as expected, non-native Arabic listeners performed more poorly than native listeners with English phonemes, and that HP listeners performed better than LP listeners, but that these differences were larger for vowels than for consonants. Indeed, HP listeners had very high recognition rates for consonants in quiet (86.5%) and in low noise conditions, performing similarly to SSBE listeners at 0dB and -5dB SNRs. Their performance in quiet was similar but not as high as reported in studies comparing Arabic-English bilinguals and American English native speakers, indicating that although our HP listeners performed well, they were not as accurate as native speakers or bilinguals (cf. Shafiro et al., 2012). LP listeners were less accurate in quiet, and performed more poorly in noise. However, as in previous studies (e.g., van Dommelen & Hazan, 2010), non-native listeners were not more adversely affected by noise vis-à-vis native speakers. Likewise, there was no reliable effect of vowel context on identification for HP & LP listeners in quiet (see also Shafiro et al., 2012) or in noise.

In terms of consonant identification, both HP & LP listeners found similar phonemes problematic; postalveolar affricates and fricatives (/tʃ, dʒ, ʒ/) dental fricatives and the velar nasal, /ŋ/, were particularly poorly identified. Poor identification of /dʒ/ was surprising, as this sound is present in MSA (Al-Ani, 1978; Amayreh and Dyson, 1998). Instead, both HP and LP participants identified this phoneme as either /g/ or /dʒ/, though HP participants reached Best and colleagues' categorization threshold of 50% (see Faris et al., 2016). As in Shafiro et al. (2012), one possibility is that low accuracy for this phoneme was driven by orthographic confusion between 'g' (/g/) and 'j' (/dʒ/), even though we had included a common word for comparison and listeners had been familiarized with the response options prior to completing the task. However, unlike Shafiro et al., (2012) we also found low accuracy for /tʃ/ for LP listeners. The majority of

errors for this group were with /ʃ/ (29%) but there were also difficulties with voicing, with LP listeners confusing this with /g/ (17%), as for /dʒ/, indicating a dispersed uncategorized response.

Likewise, errors with dental fricatives reflected voicing confusion, particularly for LP learners. MSA contains the dental fricatives /θ/ and /ð/, but as for native English listeners, Shafiro et al. (2012) found that the voiced dental fricative, /ð/, was frequently misidentified as /v/. However, our participants not only misidentified /ð/ as /v/, but also as /θ/. Indeed, for LP learners, voicing confusion accounted for more errors (/ð/ identified as /θ/; 19%) than confusion with /v/ (8%), and interestingly this error pattern also persisted in HP learners (/ð/ as /v/ 9%; /ð/ as /θ/ 9%). In contrast, /θ/ followed a more predictable pattern of assimilation; /θ/ was most commonly misidentified as /f/ (27%), but also as /ð/ (14%). One possibility is that this also reflects orthographic confusion surrounding the response options. However, this pattern of results may also reflect the influence of their L1; Arabic listeners have /f/ in their inventory but not /v/, so it is possible that they are more likely to assimilate /θ/ to /f/ than /ð/ to /v/ given the availability of this contrast in their native inventory.

Difficulties in voicing perception in LP listeners were also reflected in performance with the /p-/b/ contrast; these listeners performed well with /p/ (74%) but /b/ was less well identified (68%) and most often confused with /p/ (27%). Again, this error is not surprising given typical voicing patterns in Arabic. Note that even though the consonants identified here were produced in intervocalic position, stress was placed on the second syllable, and inspection of the waveforms revealed that our speaker produced these tokens with no voicing during the closure (cf. Docherty et al., 2011) giving a plosive with short-lag VOT. Like English, Arabic has a two-category distinction, but voiced stops are typically pre-voiced and voiceless stops have short lag (Flege, 1981; Khattab, 1999). This means that English voiced stops like those produced here, fall within the range of Arabic voiceless stops. Interestingly, HP listeners had fewer problems with voicing, particularly for the /p-b/ contrast where identification was very high (86% and 91% respectively), indicating that with experience, Arabic learners are able to learn to perceive this contrast.

As predicted, given the small vowel inventory of Saudi Arabic, performance on vowels was much poorer for non-native than native listeners, particularly LP learners; our HP learners performed similarly in quiet to the Arabic-English bilinguals tested by Shafiro et al. (2012); (69% and 70% correct respectively) but LP listeners performed more poorly overall (47%).

Although it is difficult to directly compare our results with those of Shafiro et al., (2012) as they were testing perception of American rather than SSBE vowels, there are some notable differences, in particular, for the close-mid vowel /ɪ/. In Shafiro et al., (2012), this vowel was identified with high accuracy (91%) and was rarely confused with *heed* or *head*. In our study however, LP listeners consistently identified /ɪ/ as the English mid-vowel /e/ (*hid* identified as *head* 72%), and though HP listeners showed some improvement they also frequently confused /ɪ/ with /e/ (*hid* identified as *head* 44%). In contrast, *head* was identified with a high level of accuracy by both groups (HP: 85%; LP: 69%) and rarely confused with *hid*. Likewise, HP & LP learners also found the English high back vowels /u:/, ʊ/ difficult. As for /ɪ/, this is not surprising given that Arabic has the high back vowels /u:/ and /u/, both of which are very different from the fronted English /u/ (e.g., Wells, 1982). The poor match between SSBE and Arabic /u:/ may thus explain why both HP & LP listeners performed poorly with SSBE /u:/ but not /i:/, which is a good match to Arabic /i:/, and why they frequently confused English /u:/ with /ʊ/.

Listeners also had difficulties with the central vowels, /ɜ:/ (*heard*) and /eə/ (*haired*). Interestingly, Faris et al. (2016) also found that the AusE counterparts for these vowels were uncategorized-dispersed, indicating that no Arabic vowel was a good match to these vowels. We did not collect assimilation data, but our LP participants confused these vowels with each other rather than with a number of different categories, i.e., both were identified as *heard* or *haired*, which would be more consistent with an uncategorized-clustered assimilation pattern. For HP participants, *heard* was well identified (79%) but *haired* was confused with *heard* (33%) and *head* (21%). This may indicate that these HP learners could distinguish these vowels to some extent, but that as for the /ɪ/-/e/ contrast, they were unable to do so reliably, perceiving them instead as belonging to a single category, likely with significant overlap (cf. Best & Tyler, 2007). This may also be compounded by acoustic factors; /eə/ has very little formant movement and its onset is somewhat similar to that of the central vowel /ɜ:/.

There were some similarities with Shafiro et al., (2012) however; our listeners also had problems with low-front vowels, confusing /ɒ ʌ əʊ/. At least for some of these vowels, it is possible that orthography may again have affected identification performance. For instance, /ɒ/ (*hod*) was misidentified as /əʊ/ (*hoed*) and /ʊ/ (*hood*) which are very different acoustically, but which have similar orthography and which L2 learners may have associated with the same pronunciation. It is possible then, that even though the response buttons included familiar rhyme

words, participants had not yet developed representations that were robust enough for them to be able to use this information effectively in identifying these vowels.

Unlike for consonants, noise more adversely affected native than non-native listeners' vowel identification (cf. Cutler et al., 2004). This could have been due to overall low performance, as both HP and LP learners performed more poorly even at easier SNRs. There was also no reliable evidence that Arabic listeners relied more on duration than did native listeners; all groups performed more poorly on duration-equated than natural vowels. Previous work has demonstrated that different acoustic cues correlate differently with proficiency. Iverson & Evans (2007) tested learners with different language backgrounds in their identification of natural vowels in quiet, vowels in noise that had been signal-processed to flatten formant-movement and equate duration, and also had them map best exemplars for English vowels in a five-dimensional space that included formant movement and duration. Representation of formant movement was significantly correlated with identification accuracy, but duration was only weakly correlated with identification performance in quiet, indicating that F1/F2 target frequencies contribute more to individual differences in L2 vowel identification in English, and that duration is a secondary cue that can have value in noisy conditions when the formant information is less clear (Iverson & Evans, 2007). One possible reason for why we failed to find a significant effect of duration may have been because it was restricted to individual vowel categories. For example, inspection of the data showed that HP but not LP listeners had a marked drop in performance for *hid*. LP listeners performed poorly with this vowel overall (natural 23%; duration-equated 23%) and consistently misidentified it as *head* (natural 23%; duration-equated 29%). In contrast, HP listeners performed better with this vowel in the natural condition (44%) though also tended to misidentify it as *head* (38%), but in the duration-equated condition performance dropped to 9% with incorrect responses split between *head* (30%) and *heard* (24%). These patterns are consistent with the idea that cue usage may differ according to proficiency, such that in noisy conditions higher proficiency learners may be able to make use of duration as a secondary cue. It remains for this to be investigated in future studies.

In sum, this experiment demonstrated that Saudi Arabic learners of English perform more poorly with English phonemes than native listeners, and that this is affected by overall proficiency with English. As expected, there were more errors for vowels than consonants, with

learners performing more poorly with vowels overall. The next experiment aims to investigate whether difficulties in English vowel perception are also reflected in production.

3. Experiment 2: Vowel Production

3.1 Method

3.1.1 Participants

The same SSBE participants who took part in Experiment 1, identified vowels and gave accent ratings on a short passage produced by the same SA participants tested in Experiment 1.

3.1.2 Stimuli & Apparatus

Recordings. After completing the perception tasks, the SA participants recorded the same 17 vowels that they had identified in the vowel identification task. Participants recorded three repetitions of each of the /hVd/ words in the carrier sentence *Say __ again*. To avoid list effects, stimuli were presented via PowerPoint, one word per slide. Participants also recorded a short passage, “The North Wind and the Sun” (IPA Handbook, 1999). Participants familiarized themselves with the passage before recording, and were then instructed to read at a conversational speed. The paragraph was also presented via PowerPoint. All recordings were made using a C1U USB microphone in a sound-attenuated room connected to a PC via an Edirol UA-25 processor, at a sampling rate of 44100 Hz (16-bit) samples/s.

Vowel intelligibility and accent rating experiments. Individual words were segmented and each word was saved into a separate wav file. Vowel repetitions were checked for clarity, and for each speaker the best repetition (i.e., clear voice quality, no hesitation) was chosen as the stimulus for the intelligibility task. This gave a total of 442 stimuli: 17 vowels per speaker (N=26). Stimuli were equalized for amplitude at 70 dB and downsampled to 22050 Hz.

The same accent-revealing extract was taken from each speaker's recording of the passage; “*Then the North Wind blew as hard as he could, but the more he blew, the more closely did the traveller fold his cloak around him; and at last the North Wind gave up the attempt*”. This extract was selected because it contains a range of vowels and in particular, examples of the

/ɪ/ and /e/ vowels. Extracts were saved into individual wav files, equalized for amplitude at 70 dB and downsampled to 22050 Hz.

All participants were tested in sound-attenuated rooms with stimuli played over headphones (Sennheiser HD 555) at a user-controlled comfortable level via a PC running Praat (Boersma & Weenink, 2013).

3.1.3 Procedure

Vowel intelligibility. Participants heard recordings of the SA participants' vowels in /hVd/ words in the carrier sentence “Say ___ again”, and gave a closed-set response from the 17 test words. To give their response, participants used the same interface as for the vowel identification experiment in Experiment 1; participants mouse-clicked on a button which listed the /hVd/ word and a common English word with the same vowel, (e.g., *heed* as in *seed*, *hud* as in *cut*: see Appendix B for full list). The order of the stimuli and the talker was randomized, and the identification task was self-paced with participant-controlled breaks after 50 stimuli. Participants heard each word once and received no feedback.

Accent ratings. Participants rated an extract of the SA participants' recordings of “The North Wind and the Sun”. The rating sessions were self-paced and listeners could listen to each extract twice; the order of the extracts was randomized. Listeners gave their ratings by mouse-clicking on a 7-point Likert scale where 1 was judged to be very native-like, and 7 very non-native.

3.2 Results

3.2.1. *Vowel intelligibility.* As displayed in Fig. 5, HP speakers were more intelligible (average 62% correct) than the LP speakers (average 46% correct); an independent samples t-test confirmed that intelligibility was significantly higher for the HP than the LP speakers [$t(24)=2.94, p<.05$].

To investigate whether particular vowel contrasts were difficult to produce, the data were submitted to confusion matrices. Table 5 shows the confusion matrix for the vowels produced by the LP group. *Heed* (83%) was well identified, but SSBE listeners frequently misidentified /ɪ/ (48%) as /aɪ/ (30%) or /i:/ (18%), and /e/ (28%) as /ɪ/ (36%) or /i:/ (37%). Low and central vowels were also problematic; /ɒ/ (22%) was misidentified as /ʊ/ (62%), and /ʌ/ (15%) as /ʊ/

(50%). The high back vowel /u:/ (31%) was also frequently misidentified as /ʊ/ (39%). The two diphthongs, /əʊ/ and /eə/ were particularly problematic; /əʊ/ (16%) was misidentified as /ʊ/ (35%), /ɔɪ/ (25%) or /u:/ (17%), and /eə/ (19%) was misidentified as /ɜ:/ (61%). Although overall identification rates were higher, confusion patterns for HP speakers were similar (Table 6). Listeners frequently misidentified /ɪ/ (41%) as /e/ (47%) and though /e/ was better identified (76%) this was also confused with /ɪ/ (21%). Likewise, /u:/ and /ʊ/ were confused with each other; /u:/ was frequently identified as /ʊ/ (41%) and /ʊ/ as /u:/ (45%). /ɒ/ (28%) was misidentified as /ʊ/ (47%) or /ʌ/ (21%), and /ɔ:/ (39%) was misidentified as /ɑ:/ (23%) or /aʊ/ (15%). As for LP speakers, the diphthongs /əʊ/ and /eə/ were poorly identified; /əʊ/ (34%) was misidentified as /ɔɪ/ (23%), /u:/ (16%) or /ʊ/ (16%), and /eə/ (8%) was predominantly misidentified as /ɜ:/ (62%).

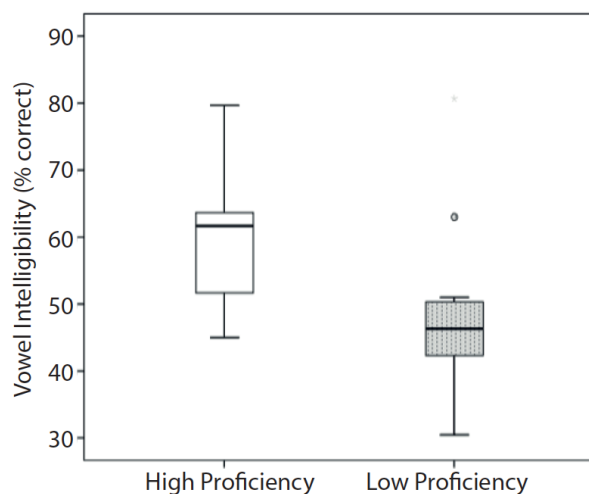


Figure 5: Boxplot showing overall vowel identification (percentage correct) of L2 speakers' productions identified by SSBE listeners. The lower and upper whiskers represent the first and last quartiles respectively, with outliers represented by shaded circles.

Table 5: The confusion matrix showing mean intelligibility (percent correct) for vowels produced by the LP group. Stimuli are in rows, and responses in columns.

| Stimulus | Response | | | | | | | | | | | | | | | | |
|-------------|----------|-----|------|-----|------|-----|-------|-------|------|-----|-------|-------|------|-------|------|--------|-------|
| | heed | hid | head | had | hard | hod | hoard | who'd | hood | hud | heard | hayed | hide | how'd | hoed | haired | hoyed |
| heed /i:/ | 83 | 15 | | | | | | | 1 | | | | | | 1 | | |
| hid /ɪ/ | | 48 | 18 | | | | | | | | | 3 | 30 | | | | 1 |
| head /e/ | 37 | 35 | 28 | | | | | | | | | | | | | | |
| had /æ/ | | 1 | 13 | 72 | 1 | 1 | | 1 | 1 | 5 | 3 | | | | | 1 | 1 |
| hard /ɑ:/ | | | 5 | 64 | 5 | 2 | | | | 1 | 22 | | | | | 1 | |
| hod /ɒ/ | | | | | | 22 | | 4 | 62 | 9 | | | | 3 | | | |
| hoard /ɔ:/ | | | | | 34 | 4 | 30 | | 3 | 1 | 18 | | 1 | | 7 | 2 | |
| who'd /u:/ | | | | | | 9 | 1 | 31 | 39 | | | | | 8 | 12 | | |
| hood /ʊ/ | | | | | | 5 | | 41 | 52 | 1 | | | | | 1 | | |
| hud /ʌ/ | | 9 | | 6 | 1 | 5 | | 5 | 50 | 15 | | | 6 | | 2 | 1 | |
| heard /ɜ:/ | 1 | 3 | 2 | | 1 | | | 4 | 4 | 1 | 62 | 1 | | | | 18 | 3 |
| hayed /eɪ/ | | | 3 | | | | | | | | | 30 | 64 | | | 1 | 2 |
| hide /aɪ/ | | 4 | 5 | | 1 | | | | | | | | 83 | 1 | | 6 | |
| how'd /aʊ/ | | | | | 1 | 4 | | 10 | 21 | 9 | | | | 52 | | 3 | |
| hoed /əʊ/ | | | | | | 4 | | 17 | 35 | 1 | | | 2 | | 16 | | 25 |
| haired /eə/ | | | 1 | | | | | | | 1 | 61 | 7 | 11 | | | 19 | |
| hoyed /ɔɪ/ | | | | | 1 | | | | | | | 1 | 14 | 1 | 2 | | 81 |

Separate hierarchical cluster analyses for the vowels produced by the LP and HP proficiency groups were used to analyse these confusion patterns. For the LP group, there were four distinct confusable clusters of vowels: the front vowels, including front closing diphthongs, the high back and low central vowels including high back closing diphthongs, the central vowels, and the back vowels /ɑ:/ and /ɔ:/. As expected, based on the confusion matrices, certain groups of vowels within these clusters were highly confusable; /ɜ:/-/eə/, /ɒ/-/ʌ/, /ʊ/-/əʊ/-/u:/, /eɪ/-/aɪ/ and /ɪ/-/e/. The clusters for the HP group showed similar patterns. SSBE listeners frequently confused HP Arabic speakers' high back vowels, /u:/ and /ʊ/, the central vowels /ɜ:/ and /eə/, and the high front vowels /ɪ/ and /e/. The vowels /ɒ/ and /ʌ/ were somewhat confused, as were /əʊ/ and /ɔɪ/.

Table 6: The confusion matrix showing mean vowel intelligibility (percent correct) for vowels produced by the HP group. Stimuli are in rows, and responses in columns.

| Stimulus | Response | | | | | | | | | | | | | | | | | |
|-------------|----------|-----|------|-----|------|-----|-------|-------|------|-----|-------|-------|------|-------|------|--------|-------|----|
| | heed | hid | head | had | hard | hod | hoard | who'd | hood | hud | heard | hayed | hide | how'd | hoed | haired | hoyed | |
| heed /i:/ | 96 | 4 | | | | | | | | | | | | | | | | |
| hid /ɪ/ | | 41 | 47 | | | | | | 3 | | | | 9 | | | | | |
| head /e/ | | 21 | 76 | 3 | | | | | | | | | | | | | | |
| had /æ/ | | | 5 | 73 | 2 | | | | | 2 | 17 | | | | | | 1 | |
| hard /ɑ:/ | | | 1 | 1 | 91 | 2 | | | | 1 | 1 | | 1 | 2 | | | | |
| hod /ɒ/ | | | | | 2 | 28 | | | 47 | 21 | 1 | | | | | | | 1 |
| hoard /ɔ:/ | | | | | 23 | 8 | 39 | | 1 | | 3 | | 1 | 15 | 2 | | | 8 |
| who'd /u:/ | | | | | | | | 59 | 41 | | | | | | | | | |
| hood /ʊ/ | | | | | | | | 45 | 54 | 1 | | | | | | | | |
| hud /ʌ/ | | | | 12 | 1 | 5 | | | 20 | 62 | | | | | | | | |
| heard /ɜ:/ | 1 | 1 | 2 | | | | 1 | 1 | | | 83 | | | | | | 10 | 1 |
| hayed /eɪ/ | | | | | | | | | | 2 | 2 | 61 | 33 | | | | | 2 |
| hide /aɪ/ | | 1 | | | 1 | | | | | | | 1 | 97 | | | | | |
| how'd /aʊ/ | | | | | 3 | | 3 | 3 | 3 | 3 | 2 | | | 72 | 2 | | | 9 |
| hoed /əʊ/ | | | | | | | | 16 | 16 | 1 | 1 | | 1 | 8 | 34 | | | 23 |
| haired /eə/ | | | 8 | 2 | 3 | | 2 | | | | 62 | 2 | 10 | 1 | 1 | 8 | | 1 |
| hoyed /ɔɪ/ | 1 | | | 1 | 1 | 1 | | | 1 | 1 | | 5 | 1 | 15 | | | | 73 |

3.2.2 *Accent Ratings*. Pearson correlations between all pairs of raters demonstrated that SSBE listeners' accent ratings were in the range of $r = .621$ to $.94$. Consequently, an average rating was calculated for each SA participant and these values were used in all subsequent analyses.

As displayed in Fig. 6, ratings for HP and LP learners were highly variable, though HP participants appeared to be judged to sound more native-like overall than LP participants. An initial analysis using an independent samples t-test and including all data points indicated that there was no significant difference between groups, $p > 0.05$. However, this result appeared to be driven by the existence of an outlier in the HP group (see Fig. 6) and an analysis excluding this outlier, demonstrated that there was a significant difference between groups, $[t=-2.18, p<.05, df=23]$, indicating that HP participants tended to be judged to sound more native-like than LP participants.

A Pearson correlation between the accent ratings and vowel intelligibility scores investigated whether participants who were rated as more native-like also tended to be more intelligible. Ratings and intelligibility scores were significantly correlated, $r = -.46, p < 0.05$,

$R^2=.165$, suggesting that although other factors may have also played a role (e.g., voice quality and intonation), listeners were sensitive to vowel quality when judging foreign accent.

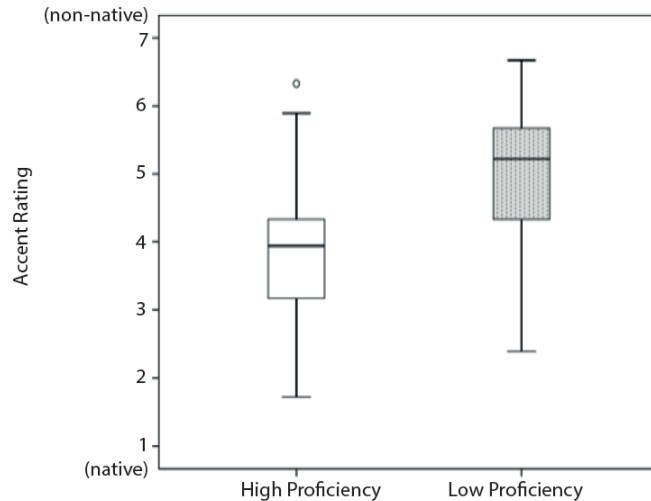


Figure 6: Boxplots showing SSBE listeners' accent ratings for L2 Arabic participants' speech. Ratings were made on a scale from 1 (native-like) to 7 (strong foreign accent); a lower rating indicates a more native-like accent. The lower and upper whiskers represent the first and last quartiles respectively, with outliers represented by shaded circles.

3.2.3 Comparison of vowel perception and vowel intelligibility. To investigate whether performance in speech perception was an indicator of ability in production, a Pearson correlation investigated the relationship between vowel identification (SSBE vowels identified by SA participants) and vowel intelligibility (vowels produced by SA participants and identified by SSBE listeners). There was a significant correlation between vowel identification and vowel intelligibility, [$r=.588$, $p<.05$, $R^2= .34$], indicating that SA participants who performed well on the vowel identification task were also more accurate in their vowel production.

Informal comparison of the confusion matrices and cluster analyses for vowel identification and vowel intelligibility suggested that groups of vowels that L2 learners found difficult to identify, were also less well identified by SSBE listeners. This was particularly noticeable for the LP group. These participants frequently misidentified /eə/ as /ɜ:/ in the vowel identification task, and their production of /eə/ was also misidentified by SSBE listeners as /ɜ:/.

The vowels /ɪ/, /e/, and the back vowels /u:/, /ʊ/, /əʊ/ were similarly confused in both perception and production.

Interestingly, perception and production were mismatched for some vowels regarding either the degree of confusion or the change in the confusion pattern. LP participants misidentified the vowels /eɪ/ (59%) and /aɪ/ (46%) but were able to differentiate between them in production (/eɪ/, 34%; /aɪ/, 84%). They performed badly with /ʊ/ in both production and perception, but in perception they misidentified this vowel as /ɑ:/ (38%) or /ʌ/ (18%) whilst their productions were predominantly misidentified as /ʊ/ (62%) rather than /ʌ/ (9%).

Similar patterns emerged for the HP group. HP participants frequently identified /eə/ (28%) as /ɜ:/ (33%), and their productions of /eə/ (8%) were misidentified by SSBE listeners as /ɜ:/ (61%). The vowel /ɪ/ was also problematic in both perception and production tasks; SA participants often identified /ɪ/ as /e/ (47%) and SSBE listeners likewise heard their production of /ɪ/ as /e/ (47%). However, whilst HP participants were able to accurately identify /e/ (84%), their production of this vowel was often confused with /ɪ/ (21%). Likewise, /u:/ was confused with /ʊ/ in perception and production, but /ʊ/ was confused with /u:/ in production but not in perception, where it was more often identified relatively reliably (68%). It is possible that even these more advanced HP learners may have been affected by the orthography and that they associated the ‘double oo’ spelling in *hood* with the longer /u:/ vowel, thus producing this word incorrectly (see also Giannokopoulou et al., 2017).

3.3 Discussion

As expected, proficiency affected production accuracy, with HP participants judged to be more intelligible than LP participants. However, accent ratings were highly variable with only a marginal difference between HP and LP participants. Although there was a weak correlation between accent ratings and vowel intelligibility, this indicates that, as might be expected, native listeners were basing their judgement in the accent rating task on other features (see General Discussion).

However, there was a clear relationship between perception and production, with those participants who performed more highly on the vowel perception task also being more intelligible. Likewise, error patterns in production often reflected difficulties in perception; all participants were least intelligible in their production of the high front vowel contrast *hid-head*,

the low central vowels *hod* and *hud*, the high-back vowel *who'd* and the diphthongs *hoed* and *haired*, all of which were problematic in perception. However, though some error patterns were the same across production and perception (e.g., for *hid-head*), this was not always the case and often differed according to proficiency (e.g., *hod* for LP learners but *hood* for HP learners). This likely reflects the instability of representations and the varying influence of orthography at different stages of the learning process.

4. General Discussion

This study provided initial information about how Saudi Arabic learners of English of varying proficiency levels perceive and produce the English phoneme inventory. We used a set of perception and production tasks to investigate the problematic phonemic contrasts for adult Saudi Arabic learners of British English. Specifically, we tested whether low (LP) and high (HP) proficiency groups had difficulty with the perception of the same phoneme contrasts, and how background noise affected the performance accuracy of both proficiency groups compared with native SSBE listeners for natural vowels and consonants and duration-equated vowels. In addition, two production tasks further investigated whether there was a relationship between perception and production accuracy for vowels, which we had hypothesized would be particularly problematic for Arabic learners of English.

The results from the perception and production tasks demonstrated consistent differences between the two proficiency groups in terms of their phoneme identification accuracy in both quiet and noise conditions. For consonants, overall performance for both groups was relatively high, but for vowels, overall performance was much lower. Reduced vowel accuracy in comparison to consonant accuracy is likely due to these listeners' L1 background, i.e., the effects of the differences in the Arabic and English phonemic inventories on L2 learning. Previous work has argued that L2 learners use their L1 categories when listening to their L2 (e.g., Best, 1995; Best et al., 2001; Flege, 1995, 2003) such that they assimilate L2 sounds into their L1 categories. Such a strategy would likely be problematic for those with small L1 inventories, because multiple L2 vowels will assimilate to the same L1 category (see Iverson & Evans, 2007) or will be uncategorizable (Faris et al., 2016). One explanation for the difficulties our Arabic listeners, in particular our LP learners, had with vowels but not consonants then, is that they were mapping the larger English vowel inventory to their smaller Arabic vowel system (Iverson &

Evans, 2007; Shafiro et al., 2012). Indeed, listeners in this study performed most poorly with vowels that did not have a direct counterpart in Arabic, i.e., were uncategorizable [/e/ (*head*), /ɜ:/ (*heard*), /eə/ (*haired*), /ɒ/ (*hod*), /ʌ/ (*hud*), /əʊ/ (*hoed*), /u:/ (*who'd*), /ʊ/ (*hood*)]. In contrast, there are more possibilities for direct one-to-one mapping across the two consonant inventories, and so even our LP learners with less experience with English were able to perform well in consonant identification.

With experience, it has been hypothesized that listeners can establish new categories and that this is easier when the new L2 categories are far away from their existing L1 categories (Flege, 1995). Our results indicate that this process starts to happen relatively early in learning and that learners likely continue to refine these categories. As previously mentioned, Arabic does not have the English /ɪ/ or /e/ vowels, but it does have the high front vowels, /i:/ and /i/, that differ in duration. Previous work has shown that learners of other languages that lack the English /i:/-/ɪ/ contrast, but which have a single high front vowel, /i/ (e.g., French, Spanish), typically assimilate both English /i:/-/ɪ/ to their native /i/ category (e.g., Iverson & Evans, 2007). However, Faris et al. (2016) found that EA listeners assimilated the contrast differently based on their L1 phonology; AusE /i:/ assimilated as uncategorized-focalized to EA /i:/ but that /ɪ/ was assimilated as uncategorized-clustered to EA /i:/ and /i/, whilst AusE /e/ was assimilated as categorized to EA /i/. In our study, both HP & LP listeners identified SSBE /i:/ and /e/ with a high degree of accuracy, but for /ɪ/, identification performance differed according to proficiency; LP listeners identified /ɪ/ almost exclusively as /e/, but HP listeners' responses were split between *hid* (46%) and *head* (44%). In SSBE and other varieties of English, /ɪ/ is typically closer to /e/ than /i/ in terms of F2 (Evans et al., 2007; Wells, 1962), and /ɪ/ and /e/ are more similar to each other in terms of duration than /i:/ is to /ɪ/ (Wells, 1962; see also Appendix B). Given the acoustic-phonetic proximity of /ɪ/ and /e/, our LP listeners may thus have initially assimilated SSBE /ɪ/ and /e/ into their native Arabic /i/ category which they mapped to English *head*. With more experience, however, learners appear to be able to differentiate the /ɪ/-/e/ contrast to some extent. Our HP listeners were thus able to unambiguously identify SSBE /i:/ and /e/, but were unsure what to do with /ɪ/ (cf. Cutler et al., 2005). This interpretation would support an account in which learning is based on phonetic rather than phonological proximity (cf. Peperkamp, 2015) but where this interacts with new category formation, which itself is dependent on the relationship between the L1 and L2 (cf. PAM-L2; Best & Tyler, 2007).

Consistent with PAM-L2 (Best & Tyler, 2007), our results also support the view that adjusting existing phonological representations is harder than acquiring an L2 sound with no direct L1 counterpart. For example, the voiceless affricate /tʃ/ has no direct counterpart in the Arabic consonant inventory and so one might expect listeners to assimilate this phoneme to their nearest native category, /ʃ/ and consequently, to identify it as English /ʃ/. In contrast, our learners did have the voiced affricate /dʒ/ and voiced postalveolar fricative /ʒ/ in their native inventory; /dʒ/ exists in MSA and is used in formal settings, and our pilot study showed that Saudi Arabic speakers use the variant /ʒ/ in their low variety in place of /dʒ/. Although HP and LP listeners assimilated English /tʃ/ to /ʃ/ (i.e., they misidentified English /tʃ/ as /ʃ/), it was notable that HP listeners performed much better with the voiceless than voiced affricate and that HP listeners performed better than LP listeners, who had similar identification scores for both /dʒ/ and /tʃ/. Further, /dʒ/ was very rarely, if at all, identified as /tʃ/. Instead, both HP and LP listeners confused /dʒ/ and /ʒ/, or identified them as the velar plosive /g/. This pattern of results suggests that although Arabic learners had difficulties acquiring the English affricate contrast, they found it easier to acquire the voiceless (no direct L1 counterpart) than the voiced affricate. Additionally, acquisition of the voiced affricate may have been affected by participants' dialect background. One possibility is that /dʒ/ (MSA) and /ʒ/ (low variant) are allophonic variants of a single underlying category in listeners' L1, and that consequently English /ʒ/ and /dʒ/ are assimilated into this single underlying native category [i.e., a single category assimilation; Best, (1994)]. This suggests that phoneme categorization may be highly specific, and that L1 dialect may play a significant role in L2 perception (see also Chladkova & Podlipsky, 2011).

As expected, accuracy of both vowel and consonant identification decreased as the noise level increased for all participants, but native listeners performed better than non-native listeners with vowels and consonants in noise (see also Cooke et al., 2008), and HP listeners outperformed LP listeners. However, noise affected the identification of vowels and consonants differently. In contrast to native listeners who performed better with vowels than consonants in noise, our Saudi Arabic listeners found vowel identification in noise harder than consonant identification. Previous work by Cutler et al. (2004) showed that Dutch listeners' identification of English vowels was not greatly affected by noise, but that identification performance for consonants was poorer in their lowest noise condition (0dB SNR). One possibility is that these differences arose because their SNRs were higher (i.e., less noise and easier to understand; 0, 8,

16 dB vs. 0, -5, -10 dB in this study). Another possibility is that because Cutler et al.'s participants were Dutch and Dutch has a more complex vowel space, their participants were able to rely on direct mapping between Dutch and English vowels. As previously discussed, Arabic listeners, even those who perform well in quiet, may not be able to rely on such strategies (i.e., being unable to map to native categories) because Arabic has a much smaller vowel space, and this may mean that they are reliant on new, less well-defined categories, which break down more easily in noise. This is consistent with our results which showed that whilst noise affected consonant identification for all subjects equally, non-native listeners' vowel identification performance was affected more at lower noise levels.

Although this study only investigated the relationship between the perception and production of vowels, there was some evidence for a link between production and perception (cf. Bradlow et al., 1997). Accent ratings and vowel intelligibility (i.e., SSBE listeners' identifications of Arabic participants' vowels) were significantly correlated; Arabic participants who were given more native-like ratings were also more intelligible. Vowel identification and vowel intelligibility were significantly correlated and there were also similarities in the error patterns in production and perception. That is, the same problematic vowel categories in perception were found to be problematic in production [e.g., /u:/ (*who'd*), /ʊ/ (*hood*), and /ɪ/ (*hid*)]. However, there were some differences and vowel categories which were not confusable in perception, were found to be confusable in production, [e.g., /ɔ:/ (*hoard*)-/ɑ:/ (*hard*)]. Furthermore, there was no correlation between vowel identification performance and accent ratings. Accent ratings were made on a short extract which was selected to include problematic vowel contrasts, in particular /ɪ/-/e/, and one possible reason for the lack of a relationship between identification and ratings, is that factors such as voice quality and prosody may have also affected ratings. That is, SSBE listeners might have found a speaker highly intelligible, but based their accent ratings on more global speech characteristics as well as, or instead of, intelligibility. Equally, it is possible that we did not include enough examples of the problematic vowel categories. In their study, Hattori and Iverson (2009) used an accent revealing sentence, "*The red robin looked across the lovely lake*", to show that Japanese learners' baseline English /ɪ/-/I/ identification was correlated with degree of /ɪ/ accent and /ɪ/-/I/ contrast. In this study, it was not clear at the outset which vowels would be problematic for Arabic learners of English and so a sentence that included vowels that were expected to be difficult, i.e., /ɪ/ and /e/, alongside

others was chosen. It is possible that as well as our ratings reflecting more global characteristics, we did not include enough instances of these problematic phonemes, which meant that they did not correlate well with overall vowel identification.

5.0 Conclusions

In conclusion, the current study is the first to explore problematic vowel and consonant contrasts for Saudi Arabic learners of English of different proficiency levels. As expected, contrasts that do not occur in Arabic presented the most difficulty for learners. In particular, Arabic learners had difficulties with English affricates, and high front, high back and central vowels. Although learners appeared to rely more on mapping non-native to their native categories when identifying consonants (cf. Iverson & Evans, 2009), with this being affected by their dialect background, there was some evidence that they had started to establish new vowel targets within their native vowel space, e.g., for the SSBE /ɪ/-/e/ contrast. This process likely starts early in learning but even highly experienced learners continue to find vowels with no direct counterpart in Saudi Arabic difficult to identify in a minimal pair task. Additionally, the study provides some evidence for a link between perception and production; perception of English vowels was better in Saudi learners who also had more accurate production of these vowels. It remains for future studies to further investigate the full extent of the potential for new perceptual category learning, and whether these categories form the basis of new articulatory targets.

Acknowledgements

This research was supported by a scholarship from King Abdul Aziz University, Saudi Arabia, to the second author.

Appendix A: Response options for Consonant identification tasks

| Consonant | Keyword |
|------------------|-----------------------|
| p | <u>P</u>ilot |
| b | <u>B</u>ear |
| m | <u>M</u>ug |
| f | <u>F</u>lower |
| v | <u>V</u>ery |
| θ | <u>T</u>heatre |
| ð | fa <u>TH</u> er |
| t | <u>T</u>oy |
| d | <u>D</u>oor |
| n | <u>N</u>ine |
| s | <u>S</u>tar |
| z | <u>Z</u>ebra |
| ʃ | <u>S</u>Harp |
| ʒ | plea <u>S</u> ure |
| tʃ | <u>C</u>Hip |
| dʒ | <u>J</u>ourney |
| g | <u>G</u>olf |
| k | <u>K</u>ey |
| ŋ | si <u>NG</u> |
| h | <u>H</u>at |
| l | <u>L</u>emon |
| r | <u>R</u>omeo |
| w | <u>W</u>ater |

Appendix B: Response options (keyword & rhyme word) and durations of natural vowels

| Vowel | Keyword | Rhyme Word | Duration (ms) |
|--------------|----------------|-------------------|----------------------|
| i: | heed | seed | 315 |
| ɪ | hid | kid | 197 |
| e | head | bed | 198 |
| æ | had | pad | 206 |
| ɑ: | hard | card | 328 |
| ɒ | hod | pod | 212 |
| ɔ: | hoard | sword | 326 |
| u: | who'd | food | 354 |
| ʊ | hood | could | 189 |
| ʌ | hud | cut | 180 |
| ɜ: | heard | bird | 345 |
| eɪ | hayed | paid | 331 |
| aɪ | hide | bite | 361 |
| ɔɪ | hoyed | toyed | 411 |
| aʊ | how'd | cowed | 387 |
| əʊ | hoed | code | 333 |
| eə | haired | paired | 383 |

Reference List

- Al-Ani, S. (1978). The development and distribution of the Qaaf in Iraq. *Readings in Arabic linguistics*. Bloomington: Indiana University Linguistics Club, 103-12.
- Allan, D. (1992). *Oxford Placement Tests 1*. Oxford University Press, Oxford, UK.
- Al-Tamimi, J. (2007). Static and dynamic cues in vowel production: A cross dialectal study in Jordanian and Moroccan Arabic. In *Proc. of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken, Germany.
- Amayreh, M. M., & Dyson, A. T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research* 41(3), 642-653.
- Bani-Yassin, R. and Owens, J. (1987). The Phonology of a Northern Jordanian Arabic Dialect. In *Zeitschrift der Deutschen Morgenlandischen Gesellschaft*, 137(2), 297-331.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words*, (pp. 167-224). Cambridge, MA: MIT Press.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, (pp. 171-204). Timonium, MD: York Press.
- Best, C.T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O-S. Bohn & M.J. Munro, *Language experience in second language speech learning: In honour of James Emil Flege*, (pp. 13-34). John Benjamins.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775 - 794.
- Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.51, retrieved 2 June 2013 from <http://www.praat.org/>

- Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, (pp. 279-304). Timonium, MD: York Press.
- Bradlow, A.R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–310.
- Chládková, K., & Podlipský, V. J. (2011). Native dialect matters: perceptual assimilation of Dutch vowels by Czech listeners. *The Journal of the Acoustical Society of America*, 130(4), EL186–92.
- Cooke, M., Lecumberri, M. G., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414-427.
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, 116(6), 3668.
- Docherty, G.J., Watt, D., Llamas, C., Hall, D. & Nycz, J. (2011). Variation in Voice Onset Time along the Scottish-English border. In the *Proceedings of the International Congress of Phonetic Sciences*, Hong Kong, 17-21 August 2011.
- Escudero, P. & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies of Second Language Acquisition*, 26: 551-585..
- Escudero, P. & Williams, D. (2011). Perceptual Assimilation of Dutch vowels by Peruvian Spanish listeners. *Journal of the Acoustical Society of America* 129(1): EL1-EL7.
- Evans, B.G., Mistry, A. & Moreiras, C. (2007). An acoustic study of first- and second-generation Gujarati immigrants in Wembley: Evidence for accent convergence? In the *Proceedings of the International Congress of Phonetic Sciences*, Saarbrücken, Germany, 6-10 August 2015.
- Faris, M., Best, C.T. & Tyler, M. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *Journal of the Acoustical Society of America* 139(1): EL1-EL5.

- Ferguson, C. A. (1959). Diglossia. *Word* 15, 325-40.
- Flege, J.E. (1981). The Phonological Basis of Foreign Accent: A Hypothesis. *TESOL Quarterly* 15 (4), 443:455
- 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*, 233-277. Timonium, MD: York Press.
- Flege, J. E. (1999). Age of learning and second language speech. Second language acquisition and the critical period hypothesis. In D. Birdsong (Ed.) *Second Language Acquisition and the Critical Period Hypothesis*, (pp. 101-132). Hillsdale, NJ: Lawrence Erlbaum.
- Flege, J. E. (2002). Interactions between the Native and Second-language Phonetic Systems. In P. Burmeister, T. Piske and A. Rohde (Eds) *An Integrated View of Language Development: Papers in Honor of Henning Wode* (pp. 217-244). Trier: Wissenschaftlicher Verlag.
- Giannokopoulou, A., Brown, H., Clayards, M. & Wonnacott, E. (2017). High or Low? Comparing High- and Low-Variability Phonetic Training in Adult and Child Second Language Learners. *Peer-J*.
- Gottfried, T. and Beddor, P. S. (1988). *Perception of spectral and temporal information in French vowels*. *Language & Speech* 31; 57–75.
- Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *The Journal of the Acoustical Society of America*, 110(1), 489-503.
- Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, 125(1), 469–79.
- Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.
- Iverson, P., & Evans, B. G. (2007). Auditory training of English vowels for first-language speakers of Spanish and German. English. *The Journal of the Acoustical Society of America* 122(5), 1625-1628.

- Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126(2), 866–77.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-B57.
- Jia, G., Strange, W., Wu, Y., Collado, J. & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America* 119(2): 1118-1130.
- Khattab G. (1999). A socio-phonetic study of English-Arabic bilingual children. *Leeds Working Papers in Linguistics & Phonetics* 7(1), 79-94.
- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30(2), 229–258.
- Morrison, G. (2002). Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results. In the *Proceedings of the North West Linguistics Conference 2002*, edited by G. S. Morrison and L. Zsoldes (Simon Fraser University Linguistics Graduate Student Association, Burnaby, BC, Canada), pp. 29–48.
- Nickerson, C & Camiciottoli, B.C. (2013). Business English as a Lingua Franca in Advertising Texts in the Arabian Gulf. *Journal of Business and Technical Communication* 27(3), 329-352
- Peperkamp, S. (2015). Phonology vs. phonetics in loanword adaptations: A reassessment of English vowels in French. In J. Romero & M. Riera (Eds) *The Phonetics-Phonology Interface: Representations and Methodologies* (pp. 71-90). John Benjamins.
- Shafiro, V., Levy, E. S., Khamis-Dakwar, R., & Kharkhurin, A. (2013). Perceptual Confusions of American-English Vowels and Consonants by Native Arabic Bilinguals. *Language and Speech*, 56(2), 145-161.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: phonetic context effects. *The Journal of the Acoustical Society of America*, 122(2), 1111–29.

- Van Dommelen, W. A., & Hazan, V. (2010). Perception of English consonants in noise by native and Norwegian listeners. *Speech Communication*, 52(11), 968-979.
- Watson, J. C. E. (2002). *The phonology and morphology of Arabic*. Oxford University Press.
- Wells, J.C. (1962). *A study of the formants of the pure vowels of British English*. Unpublished M.A. thesis, University of London.
- Wells, J.C. (1982). *Accents of English*. Cambridge University Press.