Perceptual Confusions of American-English Vowels and Consonants by Native Arabic Bilinguals

Valeriy Shafiro
Rush University Medical Center, Chicago, USA

Erika S. Levy
Teachers College, Columbia University, USA

Reem Khamis-Dakwar
Adelphi University, NY, USA

Anatoliy Kharkhurin
American University of Sharjah, United Arab Emirates

Abstract
This study investigated the perception of American-English (AE) vowels and consonants by young adults who were either (a) early Arabic-English bilinguals whose native language was Arabic or (b) native speakers of the English dialects spoken in the United Arab Emirates (UAE), where both groups were studying. In a closed-set format, participants were asked to identify 12 AE vowels presented in /hVd/ context and 20 AE consonants (C) in three vocalic contexts: /ɑCɑ/, /iCi/, and /uCu/. Both native Arabic and native English groups demonstrated high accuracy in identification of vowels (70 and 80% correct, respectively) and consonants (94 and 95% correct, respectively). For both groups, the least-accurately identified vowels were /ɑ/, /ɔ/, /æ/, while most consonant errors were found for /ð/, which was most frequently confused with /v/. However, for both groups, identification of /ð/ was vocalic-context dependent, with most errors occurring in /iCi/ context and fewest errors occurring in /uCu/ context. Lack of significant group differences suggests that speech sound identification patterns, including phonetic context effects for /ð/, were influenced more by the local English dialects than by listeners’ Arabic language background. The findings also demonstrate consistent perceptual error patterns among listeners despite considerable variation in their native and second language dialectal backgrounds.

Corresponding author:
Valeriy Shafiro, Rush University Medical Center, Communication Disorders and Sciences, 1015 AAC, 600 S. Paulina Street, Chicago, IL, 60612, USA
Email: valeriy_shafiro@rush.edu
Keywords
American-English, Arabic, cross-dialect, cross-language, speech perception

Introduction
As a result of globalization, English often serves as a lingua franca in educational, medical and business settings worldwide (Kamwangamalu, 2003), with the number of nonnative English speakers around the world beginning to outnumber native English speakers (Jenkins, 2000). In communities in which English is spoken, successful communication is critically dependent on the English comprehension abilities of individuals engaged in social interactions. Yet, even for proficient nonnative English speakers, speech intelligibility and language comprehension may be affected by a variety of factors, including the ability to accurately perceive the speech sounds of a language (see Rubin, 1994) and of different dialects of the same language (Clopper & Bradlow, 2008). This study examined the accuracy with which native speakers of Arabic living in the United Arab Emirates (UAE), where English is the lingua franca, identified American-English (AE) vowels and consonants.

Along with several countries in the Middle East, the UAE has made a concerted effort to institute widespread English language literacy among its residents (Khalaf, 2009; Lewis, 2008). Presently, English is the primary language for a considerable portion of official and informal communication in the UAE, where a large variety of English dialects are spoken, most commonly those of South Asian origin (Leonard, 2005). American English dialects are frequently used in educational, governmental, and business settings (Clarck, 2006). The presence of AE, and the accompanying need to understand it, is expected to grow in the UAE and other Middle Eastern countries in the near future as increasing numbers of American universities open satellite campuses in the Persian Gulf region and there are increasing interactions with United States government and businesses (Rupp, 2009; US–Middle East Partnership Initiative http://mepi.state.gov). However, little is presently known about the ability of native and nonnative English speakers in the region to perceive the speech sounds of AE. Three major language-related factors may be expected to play a role in the perception of AE speech sounds in the UAE and neighboring countries: (1) the influence of the listeners’ native language, (2) the influence of the English dialect(s) in which English learning and daily communications occur in the region, and (3) the influence of the phonetic context in which individual speech sounds are presented. These factors and their potential effects on the perceptual confusions of AE speech sounds are considered separately below.

1.1 Native language effects
First, the native language of the majority of individuals in the region is Arabic. Thus, native Arabic phonology may influence their perception of nonnative English speech sounds in systematic ways, related to the relative composition of the Arabic and English phonological inventories. The effects of Arabic as a native language on the perception of AE speech sounds may be influenced by a wide variety of Arabic dialects spoken in the region. Arabic-speaking UAE residents are native speakers of Arabic dialects specific to their region of origin, but are also typically educated in Modern Standard Arabic (MSA) – a phenomenon known as diglossia (Ferguson, 1959). According to Ferguson, Arabic diglossia is characterized by the existence of the “low” and “high” language varieties of Arabic. The low language variety refers to the spoken dialects used for daily communication. The high language variety refers to MSA or Classical Arabic, which is used mainly in written texts and formal settings and unifies the Arabic-speaking world (Ferguson, 1959; Holes, 2004). Linguistic
differences between the colloquial varieties and MSA are manifested in all language domains (Abu-Rabia, Share, & Mansour, 2003; Ibrahim, 1983; Khamis-Dakwar, Froud, & Gordon, 2012; Rosenhouse, 1997). However, the separateness of these two linguistic systems is still debated in the literature (Hary, 1996; Khamis-Dakwar & Froud, 2007).

At the most general level, the phonological inventory of the high Arabic language variety, MSA, which is generically less variable among speakers of different dialects of Arabic, includes 28 consonants and six vowels. The vowels are characterized as three short and three long vowels (/i/, /i:/, /a/, /a:/, /u/, /u:/), with vowel duration being phonemically distinctive (Holes, 2004). On the other hand, the phonological inventories of the low language varieties are highly variable among Arabic speakers. Different dialects have a different number and quality of vowels and consonants. For instance, while Kuwaiti Arabic has an eight-vowel system (/i/, /i:/, /a/, /a:/, /u/, /u:/, /e:/, and /o:/), Syrian Arabic has an eleven-vowel system (/i/, /i:/, /e/, /e:/, /a/, /a:/, /u/, /u:/, /o/, /o:/, and /ə/) (Cowell, 1964). Even though all versions of colloquial Arabic’s phonological system differ from that of MSA (Ibrahim, 1986), Ferguson (1959) maintains that the sound systems of MSA and colloquial Arabic varieties constitute a single phonological system of which the colloquial phonology is the basic system and the divergent features of MSA phonology are a subsystem or parasystem. Because the two systems together have a smaller vowel inventory than the AE vowel inventory, which contains 12 stressed vowels (including /ɜ/), it may be expected that more than one AE vowel may be assimilated to a single Arabic vowel category (Best, 1995).

Consonants in all Arabic phonological inventories are marked by the presence of emphatic and pharyngeal consonants. Emphatic phones, such as /ṭ/, /ḍ/, /ð/, /ẓ/, and /ṣ/, are coronal obstruents characterized by a secondary articulation involving retraction of the tongue, resulting in a narrowing in the upper portion of the pharynx, along with their primary coronal articulation (Ghazali, 1977). Acoustic correlates of emphatic phones include a raised third formant, a lowered second formant, and a raised first formant suggesting a pattern of narrowing near the epiglottis (Jongman, Herd, & Al-Masri, 2007). Two pharyngeal consonants exist in Arabic: the voiceless /ħ/ and voiced /ʕ/ and three uvular (/x/, /γ/, and /q/) consonants. /ħ/ and /ʕ/ involve a low pharyngeal constriction and are described as approximants. /ʕ/ has vowel-like formant structures. On the other hand, emphatic and pharyngeal consonants are not found in English. It is also worth noting that the following English consonants are not found in Arabic: /p/, /v/, /ɹ/, /ʒ/, /g/, /ð/, and /ŋ/ (Amayreh & Dyson, 1998) and, thus, may not be easily mapped onto Arabic consonant counterparts.

1.2 Effects of regional English dialects

A second factor that may affect the perception of AE speech sounds in the Gulf region by L2 English speakers whose native language is Arabic involves the variety of English dialects that constitute the learning and communication environment. As previously mentioned, the majority of locally-spoken English dialects are primarily of South Asian origin (e.g., Pakistan, India, Sri Lanka, Bangladesh, Philippines), often spoken by nonnative English speakers, with highly diverse native language backgrounds. Thus, the phonological characteristics of locally-spoken English dialects that are commonly used in education, business, and social settings may also affect the perception of AE speech sounds, even among early bilinguals or native English speakers in the UAE. Previous studies indicate that English dialects can differ in their overall intelligibility for the same listener group and that perception of speech in an unfamiliar dialect or with an unfamiliar foreign accent may decrease intelligibility (Bradlow & Bent, 2003, 2008; Clarke, 2003; Clopper & Bradlow, 2008). Hypotheses regarding the possible perceptual consequences as a function of the many English dialects spoken by the listeners in the UAE would be beyond the
scope of this paper and require a great deal of preliminary research. However, as many of the English dialects stemming from East Asia are more similar to Standard British English (i.e., Received Pronunciation) than Standard American English, it may be expected that listeners in the UAE may be more familiar with this dialect, as well as other characteristics of their particular English dialect. Thus, for example, perception of AE by individuals whose English is of Philippine origin may be affected by British English vowel mergers, as well as by bilabial stop vs. labiodental fricative confusions that result from the absence of labiodental fricatives in Tagalog (Espinosa, 1997).

1.3 Phonetic context effects

Third, speech sound identification can be influenced by the phonetic context. Significant consonantal-context effects have been documented in cross-language vowel perception and production (Gottfried, 1984; Levy, 2009a, 2009b; Levy & Law, 2010; Levy & Strange, 2008). There is also some evidence of vocalic context effects on consonants in native listeners with typical speech communication (Monnin, Loevenbruck, Beckman, & Edwards, 2011; Yeon, Wayland, Harnsberger, & Silver, 2004) and by listeners with hearing impairments who use cochlear implants (Donaldson & Kreft, 2006). Effects of vocalic context on consonant identification have been also shown in at least one study for identification of nonnative Arabic consonants by Korean-Arabic bilinguals whose native language was Korean (Hong & Sarmah, 2009).

1.4 The present study

The main purpose of the present study was to provide an initial assessment of the abilities of English speakers from a linguistically-diverse Middle Eastern country to identify AE vowels and consonants. Because no previous research has addressed this question for English speakers in the Gulf region, this study took an exploratory approach that would provide general perceptual characteristics of a representative sample of young adult English speakers. Participants included (1) a group of early Arabic-English bilinguals and (2) a group of native English speakers of various geographic and linguistic origins, with all participants enrolled in a UAE university in which English is the primary language of instruction. Although the Arabic dialectal backgrounds of the participants in the first group, as well as the English dialectal backgrounds of the participants in the second group, were quite heterogeneous, as described below, these participants provided an ecologically valid representative sample of young adult English speakers, characteristic of that region. For these students, the ability to communicate in English, in general, and to perceive the sounds of AE, in particular, is expected to be most relevant at the present time and, possibly, in the future. The specific objectives of the study were to (1) evaluate the overall identification accuracy of AE vowels and consonants in proficient young adult Arabic-English bilinguals living in the UAE; (2) assess relative perceptual difficulty and classification error patterns for specific vowels and consonants; (3) examine potential effects of different vocalic contexts on consonant perception in light of previous findings demonstrating phonetic context effects on the perception of second language speech sounds (Gottfried, 1984; Hong & Sarmah, 2009; Levy, 2009a, 2009b).

It was expected that participants in both groups would demonstrate high overall accuracy in the identification of AE speech sounds because of the high overall English language proficiency required for college education. Owing to the exploratory nature of the study and little existing information about the phonology of local Arabic and English dialects, no strong predictions could be made about specific vowel and consonant error patterns or effects of phonetic contexts. However, “new” sounds (i.e., speech sounds with no phonetic counterpart in the native language) were
expected to be more accurately perceived and produced than “similar” sounds with L1 counterparts (Flege, 1995).

Thus, the following hypotheses were cautiously offered. To a first approximation, it was expected that UAE Arabic-English bilinguals would experience greatest difficulties with the AE speech sounds that have similar, but not identical, counterparts in their L1 phonological inventory than those that do not have an obvious unique phonological counterpart (Flege, 1987, 1995). This was expected to be true for the consonants /p/, /v/, /ɹ/, /ʒ/, /ð/, /g/, and /ŋ/, which comprise part of the AE, but not MSA or most of the Arabic phonological inventories. However, the magnitude of native language effects could be substantially mediated by the age of L2 learning (Best & Tyler, 2007; Flege, 1995).

Regarding vowels, given the generally smaller vowel inventory of MSA and regional Arabic dialects in comparison to the AE vowel inventory, which contains 12 stressed vowels (including /ɝ/), it was expected that more than one AE vowel would be assimilated to a single Arabic vowel category (Best, 1995). This mismatch would, therefore, result in poor identification of vowels such as /a/, /e/, /ae/, /ɛ/ with no identical Arabic counterpart. However, it was considered possible that these L2 vowels would be perceived as “new” or “uncategorized” vowels and thus with greater accuracy, but less stability, than vowels assimilated to native categories (Best & Tyler, 2007; Flege, 1995). As durational differences are contrastive in the listeners’ native Arabic (and generally more acoustically salient cues than are spectral differences), native speakers of Arabic would likely rely on durational cues (rather than spectral cues) more than would native speakers of AE (Flege, Bohn, & Jang, 1997), thus discriminating vowels differently from, and probably less accurately than native listeners.

Given the large degree of heterogeneity in participants’ native Arabic and local English dialects that might affect their perception of AE sounds, it was important to establish first whether stable error patterns would be obtained across participants in either group. It was expected that stable identification error patterns would be obtained for each group, given the early age of English acquisition for Arabic-native bilinguals and a long history of exposure to the same set of English dialects in the UAE for most participants. Cross-group comparisons could then further indicate the extent to which listener performance was driven by the participants’ native language versus local English dialect(s), thus providing a basis for future investigations. Similar error patterns between native Arabic and native English groups would indicate the dominant role of the local English dialect, while differences in identification error between the groups could suggest the effects of native language phonology.

2 Method

Vowel and consonant tests were administered to listeners to determine their overall and speech sound-specific identification accuracy. The consonant test required closed-set identification of 20 AE consonants presented separately in three different vocalic contexts: /ɑCa/, /iCi/, /uCu/. The vowel test required closed-set identification of 12 AE vowels presented in a fixed /hVd/ context. All participants were tested simultaneously in a quiet room equipped with personal computers, in two separate sessions that lasted approximately 30 minutes each. Stimuli were presented over headphones at a comfortable listening level.

2.1 Participants

Forty-two proficient English speakers enrolled in undergraduate studies at the American University of Sharjah, where English is the main language of instruction, participated. Their average age was
20 years (range 17–24), and none of the participants reported history of speech or hearing impairments. Twenty-five bilinguals (20 females, 5 males) comprised the native Arabic language group (NA), and 17 (12 females, 5 males) comprised the native English language group (NE). All participants received partial course credit for participation in an experiment.

For the 25 bilinguals in the NA group, participants’ reported places of origin were UAE (5), Saudi Arabia (5), Oman (3), Lebanon (3), Egypt (3), Bahrain (2), Palestine (2), Syria (1), and Iraq (1). Their mean age of arrival to UAE was 11 years (range 0–20), and average duration of residence in the UAE was 10 years (2–20). All considered Arabic their native language. Their average age of starting to learn English (either at home, at school or both) was 3 years, range 0–7). Most subjects considered themselves fully bilingual or more proficient in English than in Arabic and reported approximately equal use of Arabic and English in daily life. All participants also reported that they were proficient in MSA.

For the 17 participants in the NE group, reported places of origin included India (4), Pakistan (4), Sudan (2), Iran (1), Iraq (1), Malawi (1), Palestine (1), Poland (1), Qatar (1), and UAE (1). Their mean age of arrival to UAE was nine years (range 0–20), and mean duration of residence in UAE was 11 years (range 1–22).

All participants in the NE group also had proficiency in at least one additional language, which, for the majority was Urdu (6), followed by Arabic (4), with other languages being Hindi, Gujarati, Malayalam, Polish, Persian, Portuguese, French, and Japanese. Of the 17 participants, 10 had some proficiency in three of these languages, and six in four of these languages.1

2.2 Stimuli and procedure

The vowel test was based on the materials developed by Hillenbrand, Getty, Clark, and Wheeler (1995), available from http://homepages.wmich.edu/~hillenbr/voweldata.html. They included 12 AE vowels spoken in hVd contexts: heed (/i/), hid (/ɪ/), hayed (/e/), head (/ɛ/), had (/æ/), hod (/ɑ̊/), hawed (/ɔ̊/), hoed (/o/), hood (/ʊ/), who'd (/u/), hud (/ʌ/), heard (/ɝ/). Each vowel was spoken by five male and five female talkers for a total of 120 stimuli, ensuring interspeaker variability.

Vowel productions were selected from talkers who demonstrated maximum accuracy overall and for specific vowel contrasts, such as the /ɑ̊-ɔ̊/ distinction found in only some AE dialects (Strange et al., 2007).

The consonant test was based on the materials developed and provided by Shannon, Jensvold, Padilla, Robert, and Wang (1999) and included 20 AE consonants (/p, t, k, f, ð, s, ʃ, tʃ, b, d, g, v, z, dʒ, r, l, w, j, m, n/), each spoken in three vocalic contexts: /ɑ̊Ca/, /iCi/, /uCu/. Each consonant was spoken by one male and two female talkers, totaling 60 stimuli. All stimuli were presented twice in random order, resulting in 120 trials per vowel context or 360 total consonant trials.

Vowel and consonant files, sampled at 32,000 Hz and 44,100 Hz, respectively, were normalized in root-mean-square energy and presented to participants at a comfortable listening level using personal computers with Praat software (Boersma & Weenink, 2007). Both tests were given in a closed-set format and were self-paced. The consonant test was further blocked by vocalic context. In each trial, the participant heard a stimulus and selected one of the displayed response options. These were the 12 /hVd/ words with target vowels or 20 letter symbols, one for each consonant, following typical orthographic sound-to-letter conventions, for example, ‘k’ for /k/, ‘sh’ for /ʃ/ and so on. No feedback or additional familiarization was provided during the tests. During the first testing session, all participants completed the vowel test and the consonant test in one vocalic
context (/\alpha Ca/). During the second session, which took place two weeks after the first one, the participants completed the consonant test in the two remaining vocalic contexts.

3 Results

Participant responses were analyzed separately within each group before being compared across groups. Overall identification accuracy was obtained across all vowels and across all consonants, as well as separately for each of the three consonantal contexts. Next, mean speech sound identification accuracy was computed for each individual vowel and for each individual consonant in every context. Finally, confusion errors were computed for each speech sound and combined into confusion tables (Tables 1–4).

3.1 Vowels

Overall vowel identification was 70% correct for the NA group and 80% correct for the NE group, with large variation across individual vowels. A repeated measures ANOVA with 12 vowels as different levels of a single within subject variable and participant group as a between subject variable confirmed a significant main effect of vowels \( F(11, 440) = 58.44, p < 0.05 \), but no significant effect of language group or vowel by language group interaction. Low and back vowels presented the greatest perceptual difficulty for both language groups (for the NA group /\alpha/, /\o/, /\ae/ – 15, 37, 55% correct, respectively, and for the NE group /\alpha/, /\o/, /\ae/ – 16, 56, 73% correct, respectively), see Tables 1 and 2. Post hoc pairwise comparisons of individual vowel accuracy between groups revealed that vowel identification accuracy differed significantly at \( p < 0.05 \) for only a single vowel (/o/). In general, the pattern of vowel confusions in both groups suggested that acoustic and articulatory distances were not always a reliable predictor of vowel confusions for either group. For instance, the low back vowel /\alpha/, which was the least accurately perceived vowel, was more

Table 1. Vowel identification responses of the Arabic-native participants as percentages of the total for each vowel category presented. Vowel stimuli are listed in the first column and vowel response labels are listed in the top row.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>heed</th>
<th>hid</th>
<th>hayed</th>
<th>head</th>
<th>had</th>
<th>hod</th>
<th>hawed</th>
<th>hoed</th>
<th>hood</th>
<th>who’d</th>
<th>hud</th>
<th>heard</th>
</tr>
</thead>
<tbody>
<tr>
<td>heed (i)</td>
<td>88.8</td>
<td>5.2</td>
<td>1.2</td>
<td>2.8</td>
<td>0.4</td>
<td>0.8</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>hid (i)</td>
<td>2.4</td>
<td>91</td>
<td>1.2</td>
<td>2.8</td>
<td>0.4</td>
<td>0.8</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>hayed (e)</td>
<td>0.8</td>
<td>1.6</td>
<td>93.2</td>
<td>1.6</td>
<td>1.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>head (e)</td>
<td>3.2</td>
<td>2.4</td>
<td>1.2</td>
<td>80.4</td>
<td>7.2</td>
<td>1.2</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>had (æ)</td>
<td>2.4</td>
<td>0.8</td>
<td>1.2</td>
<td>38.4</td>
<td>55</td>
<td>0.4</td>
<td>0.8</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>hod (u)</td>
<td>0.4</td>
<td>1.6</td>
<td>1.2</td>
<td>34</td>
<td>15</td>
<td>10.8</td>
<td>3.6</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>hawed (ɔ)</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>23</td>
<td>23</td>
<td>37.2</td>
<td>6.4</td>
<td>2.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>hoed (o)</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>3.6</td>
<td>18</td>
<td>69.2</td>
<td>6</td>
<td>1.2</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>hood (u)</td>
<td>0.4</td>
<td>10</td>
<td>0.4</td>
<td>4.4</td>
<td>4.4</td>
<td>76.4</td>
<td>4</td>
<td>3.6</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>who’d (u)</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>6.4</td>
<td>6.4</td>
<td>22.4</td>
<td>67.6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>hud (ʌ)</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>2</td>
<td>9.6</td>
<td>2.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>heard (ɜ)</td>
<td>0.4</td>
<td>1.2</td>
<td>1.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td></td>
</tr>
</tbody>
</table>
often confused with /æ/, a low-mid front vowel than with /ʌ/, which is a more mid-central vowel, closer to /a/ in acoustic and articulatory space. On the other hand, in both groups, high front vowels were more identifiable and less confusable with other vowels despite similar acoustic distances (i.e., /e/, /ɪ/, /i/ — 93, 91, 89% correct, respectively, for the NA and 94, 98, 92% correct, respectively, for the NE).

### 3.2 Consonants

Overall consonant identification accuracy was high in both groups across vocalic contexts (i.e., 95% correct for NA and 94% correct for NE) and within each context (i.e., /aCa/, /iCi/, /uCu/ — 97, 93, 95% correct, respectively, for NA and 95, 92, 94% correct, respectively, for NE). As also found for native AE listeners in the USA (Shannon et al., 1999), both the native Arabic and native English listeners in the UAE had the highest number of errors for /ð/, which was most often mis-identified as /v/. For both the native Arabic and native English listeners in the UAE identification of /ð/ was context dependent. While for both groups /ð/ identification was poor in /aCa/ and /iCi/ contexts (i.e., 64 and 37% correct, respectively, for NA and 60 and 44% correct respectively, for NE), it was considerably more accurate in /uCu/ context (i.e., 96% and 86% for NA and NE, respectively). Between group t-tests with Bonferroni correction for /ð/ in each context confirmed the absence of significant differences for /ð/. Nevertheless, the somewhat lower /ð/ accuracy of the participants in the NE group in the /uCu/ context may be related to a number of /z/ error responses (12%) to /ð/ in the NE group only. The /z/ response to /ð/ was completely absent in the NA group for this context (Tables 3 and 4).

Additional minor but consistent confusions were found in both NA and NE groups for the consonant /r/, which was sometimes identified as /v/. Strikingly, the tendency to provide a /v/ response to /r/ was only present in /uru/ context (i.e., 11% /v/ responses for NA and 20% /v/ responses for NE). The /r-v/ confusions may result from the coarticulatory lip rounding during the production of /uru/, which could obscure the acoustic differences between the two consonants.
In contrast, NE participants alone tended to occasionally provide /v/ responses for /w/ stimuli in all three vocalic contexts (i.e., 18% for /awa/, 20% for /iwi/ and 27% for /uwu/), while NA participants gave only 3% of /v/ responses to /uwu/ stimuli and had no /v/ responses to /w/ in the two other contexts. Closer examination revealed that the seven listeners that provided /v/ responses to /w/ were primarily second language speakers of Urdu and Hindi as well as one speaker of Persian and one of Malayalam. No /v/ responses to /w/ stimuli were found among the four Arabic second language speakers or three remaining speakers of French, Polish and Portuguese (one each). This distribution of errors may suggest that the /v-w/ confusions were due primarily to the effects of specific English dialects, which might have been influenced by the phonologies of other locally-spoken languages such as Urdu and Hindi (Masica, 1991; Ohala, 1999). On the other hand, these confusions are absent in other native speakers of English dialects who might have been affected by other locally-spoken languages such as Arabic.

Finally, the unexpectedly low accuracy in responses to affricate consonant /dʒ/ demonstrated by both groups was most likely due to the ambiguity between orthographic symbols ‘g’ and ‘j’ that subjects used to indicate their responses. Because both letter names contain this affricate consonant subjects were uncertain which one constituted the correct response and responded ‘j’ 72.7% and 76% of the time and ‘g’ 27.1% and 22.0% of the time for NA and NE groups, respectively. It can be seen that if both responses were considered correct, identification accuracy for /dʒ/ would be close to 100%.

Table 3. Consonant identification responses of the Arabic-native participants as percentages of the total for each consonant category presented. Consonant stimuli are listed in the first column and consonant response labels are listed in the top row.

| Stimulus Response | p | t | k | f | th | s | sh | ch | b | d | g | v | z | j | r | l | w | y | m | n |
|-------------------|---|---|---|---|----|---|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p                 | 99.8 | 0.2 |   |   |    |   |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| t                 | 99.1 | 0.2 | 0.2 | 0.2 |    |   |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| k                 | 99.1 | 0.2 | 0.2 | 0.2 |    |   |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| f                 | 99.1 | 0.9 |    |    |    |   |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| s                 | 0.2  | 98.7 | 0.2 | 0.4 |    |   |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| tʃ                | 0.2  | 1.6 | 96.9 | 0.9 |    |   |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| b                 | 1.6  |    | 97.3 | 0.7 | 0.2 | 0.2 |    |    |   |   |   |   |   |   |   |   |   |   |   |   |   |
| d                 | 0.2  | 99.6 |    |    |    |   |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| g                 | 0.4  |    | 98.2 | 1.3 |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| v                 | 3.1  | 96.4 | 0.2 |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| z                 | 0.2  |    | 99.8 |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| dʒ                | 0.2  | 0.2 | 27.1 | 72.7 |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| r                 | 3.6  | 95.8 | 0.7 |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| l                 | 2    | 0.2 |    | 0.2 | 97.3 |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| w                 | 1.1  | 0.2 | 96.9 | 1.8 |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| j                 | 0.2  | 0.2 |    | 4.7 | 94.9 |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| m                 | 0.2  |    |    | 99.3 | 0.4 |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
| n                 | 0.2  |    |    | 1.8 | 97.8 |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
4 Discussion

4.1 Overall accuracy

This study provided initial information about the accuracy with which university students with diverse language backgrounds in a culturally and linguistically heterogeneous country of the Persian Gulf perceive AE speech sounds. In response to the first research question, regarding overall perceptual accuracy, the data suggest that these listeners perceive AE speech sounds quite accurately, regardless of their native language (i.e., 70% correct for vowels and 95% correct for consonants for the NA group and 80% correct for vowels and 94% correct for consonants for the NE group). This result was expected for both groups, given that participants in the NA group were early English bilinguals and that both groups were enrolled in university classes conducted in English. On the other hand, reduced vowel accuracy of the current NE speakers of the regional English dialects compared to native AE speakers suggests possible effects of differences in vowel inventories of various English dialects. That vowel perception was less accurate than consonant perception may be due to mapping of the larger AE vowel inventory onto the smaller vowel inventories of the regional English dialect or native Arabic, as has been shown with other language pairs (e.g., Escudero & Boersma, 2002, but cf. Flege, 1987, 1995), whereas one-to-one mapping was more likely across consonant inventories. Reduced identification accuracy could have been also due to the influence of locally-spoken English dialects, which may also differ considerably from AE.

Table 4. Consonant identification responses of the native English participants as percentages of the total for each consonant category presented. Consonant stimuli are listed in the first column and consonant response labels are listed in the top row.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>99.6 0.4</td>
</tr>
<tr>
<td>t</td>
<td>98.4 0.8</td>
</tr>
<tr>
<td>k</td>
<td>100</td>
</tr>
<tr>
<td>f</td>
<td>99.6 0.4</td>
</tr>
<tr>
<td>s</td>
<td>98.9 0.4</td>
</tr>
<tr>
<td>j</td>
<td>2 95.9 1.6</td>
</tr>
<tr>
<td>tʃ</td>
<td>0.4 99.6</td>
</tr>
<tr>
<td>b</td>
<td>0.8 98.4 0.4</td>
</tr>
<tr>
<td>d</td>
<td>2.4 97.6</td>
</tr>
<tr>
<td>g</td>
<td>100</td>
</tr>
<tr>
<td>v</td>
<td>0.4 3.3 90.2 0.4 5.3 0.4</td>
</tr>
<tr>
<td>z</td>
<td>0.4 99.6</td>
</tr>
<tr>
<td>dʒ</td>
<td>0.4 22.8 0.8 76</td>
</tr>
<tr>
<td>r</td>
<td>7.3 91.5 1.2</td>
</tr>
<tr>
<td>l</td>
<td>0.4 99.6</td>
</tr>
<tr>
<td>w</td>
<td>0.4 0.4 22.4 76.8</td>
</tr>
<tr>
<td>j</td>
<td>0.4 0.4 22.4 76.8</td>
</tr>
<tr>
<td>m</td>
<td>98.8 0.4</td>
</tr>
<tr>
<td>n</td>
<td>99.2 0.8 1.6 98</td>
</tr>
</tbody>
</table>

Downloaded from las.sagepub.com at RUSH UNIV on June 8, 2012
The overall similar performance of the native Arabic and the native English groups suggests that vowel and consonant identification errors were influenced more by the regional English dialect than native language phonology. Although there was a small, but significant, advantage in vowel identification accuracy for the NE group, identification error patterns remained largely the same as for the NA group, suggesting that the differences between the two groups were quantitative, not qualitative. The single exception was the /v-w/ distinction, which was more difficult for the NE listeners, suggesting a second language influence between the groups, as these responses were common for NE listeners whose second languages (e.g., Hindi, Urdu) did not have the /v-w/ distinction present in AE (Masica, 1991; Ohala, 1999).

4.2 Types of errors

The second research question addressed the classification errors that would occur. The error patterns that were still present for vowel and consonant identification were remarkably consistent between the two groups despite the substantial variability in participants’ native Arabic dialects and English dialectal backgrounds. Such consistency in responses suggests that participants’ perceptual classification of AE vowels and consonants was largely driven by a similar set of factors. Also of interest for the NA group was the unexpected near absence of confusions between /p/ and /b/ despite the absence of /p/ in Arabic, a finding not consistent with /p-b/ confusions by native Arabic speakers reported in the literature and Arabic speakers’ non-native-like productions of /p/, /t/, /k/ in English (Flege & Port, 1981). This finding may be attributable to the higher proficiency or earlier acquisition of English in these participants than those in previous studies or to the more accurate production of “new” speech sounds than speech sounds that are perceived as similar to native phonological categories (Flege, 1995). On the other hand, most of the consonant and vowel confusions in our study suggest that speech sounds with no counterpart in the native language/dialect were relatively difficult to identify accurately. For example, the common finding of /ð-v/ confusions and confusions involving /ɹ/ may have been expected from native speakers of Arabic, as /ð/, /v/, and /ɹ/ are included in the Arabic phonological inventory. The prevalence of similar confusions among native English speakers in both the UAE and the US, as described below, also points to the influence of locally-spoken English dialects, primarily of South Asian origin.

Generally, as expected, AE vowels with no counterpart in MSA, /ɑ/, /ɔ/, /æ/, accounted for the most vowel confusions in both groups. Listeners heard /a/ in ‘hod’ most often as /æ/ in ‘had’ followed closely by confusions with /ʌ/ in ‘hud’. For NA listeners, this may reflect the mapping of the more numerous mid and low phones of AE onto the single low /a/ category in Arabic. The reasons for this confusion are less clear in the NE listeners and might indicate a reduced vowel inventory of local English dialects compared to AE. In both groups, the /ɔ/ in ‘hawed’ was most often confused with /a/ in ‘had’. The /ɔ-a/ distinction is one made only in parts of the United States, such as the New York regional area (Strange et al., 2007). Thus this distinction might not be produced or perceived by many native speakers of English and L2 learners raised both in and out of the United States. The vowels in ‘head’ and ‘had’ were also confused frequently, findings similar to those reported by Bohn and Flege (1990) on /ɛ-æ/ confusions for German listeners. This confusion may reflect the phonological proximity of these vowels in AE and the lack of distinct Arabic counterparts. That is, English words with sparse phonological neighborhoods (i.e., smaller number of words that sound similar to each other) are recognized more accurately and rapidly than words with dense neighborhoods (Luce & Pisoni, 1998). The neighborhood spread (i.e., the number of speech sounds that may be changed to create words) also influences word recognition (Vitevitch, 2007). Thus, in terms of overall speech comprehension, it may be expected that vowel confusions associated with large numbers of
minimal pair contrasts in English might impede overall comprehension (i.e., /ɛ-æ/, /ɑ-ʌ/, /ɑ-æ/), while less frequently occurring contrasts, such as /u-ʊ/ may be less crucial (Dewey, 1923).

4.3 Vocalic context effects

In response to the third research question investigating vocalic context effects, an unexpected vocalic context effect was revealed in the present study: listeners produced /ð-v/ confusions in /ɑCɑ/ and /iCi/ context more often than in /uCu/ context. Monolinguals’ perception of their native AE consonants has shown vocalic context effects for listeners with normal hearing (e.g., Nittrouer & Stoddert-Kennedy, 1987; Whalen, 1981) and listeners with cochlear implants (Donaldson & Kreft, 2006). However, with few exceptions (Hong & Sarmah, 2009; Monnin et al., 2011) and aside from the context-related mapping patterns for /ɾ-I/ by Japanese learners of English (Lotto, Sato, & Diehl, 2004), vocalic context effects are little discussed in cross-language literature. Shannon et al. (1999) reported /ð-v/ confusions by native AE speakers, although context-specific data were not available.

Our preliminary follow-up consonant tests with six native AE listeners who were young adult university students tested with the consonant stimuli from the present study, under comparable test conditions, were consistent with Shannon et al.’s (1999) data in terms of overall accuracy. Remarkably, native AE listeners, similarly to NA and NE listeners, demonstrated numerous /ð-v/ confusions in /i/ and /ɑ/ contexts, and no /ð/ errors in the /u/ context. As with NE and NA listeners, /ð/ errors were not limited to any one talker and, in the /u/ and /i/ contexts, were present in the stimuli produced by all three talkers. Additional follow-up listening by three phonetically-trained experts to each /ð/ stimulus used in the present study as well as those produced by seven other talkers who were included in Shannon et al.’s (1999) data set revealed no recording artifacts such as noise.

For NA listeners the vocalic context effect might also be attributed to the phonemic distinction between emphatic /ð/ and non-emphatic (also interdental) /ẓ/ in Arabic dialects, which was spoken by all of the participants. Emphatic consonants are produced with an additional retraction of the tongue body (Ghazali, 1977). Thus, listeners may have assimilated AE /ð/ to their native (emphatic) consonant category in back /uCu/ context, whereas /ð/ in other contexts may have been assimilated to non-emphatic /ẓ/, resulting in different identification accuracy in the different contexts. However, data from their NE peers and AE listeners showed the same context-specific effect, suggesting a non-Arabic-specific and more general context effect for /ð-v/ confusions, possibly a characteristic of dominant local English dialects. Context-dependent variation has also been reported for individuals with and without communication disorders (Donaldson & Kreft, 2006; Hong & Sarmah, 2009; Monnin et al., 2011; Yeon et al., 2004). In Donaldson and Kreft’s study of native AE listeners with cochlear implants, /ð/ was also most accurately perceived in /u/ context, although accuracy differences from the /ɑ/ and /i/ contexts were not as large. For the cochlear implant listeners, /v/ was also the most frequent confusion error in response to /ð/ for the /ɑ/ and /i/ context, but not in the /u/ context, in which /ẓ/ and /d/ were more frequent.

Overall, these vocalic context-dependent /ð/ confusions are consistent with the coarticular effects of lip rounding on the fricative noise associated with /ð/ and /v/. Acoustic differences in the fricative noise between /ð/ and /v/ may be more clearly marked in the /u/ context due to the lip protrusion, which increases the length of the resonating cavity in front of the noise source. This possibility was tested through acoustic analysis conducted on the /ð/ and /v/ fricative noise produced by each of the talkers of the present study. Measurements were performed on the intervocalic noise portions and included noise duration, mean intensity, center of spectral gravity and its standard deviation, skewness and kurtosis. Averaged across vocalic context for each consonant, the
analysis revealed that compared to /ð/, /v/ noise had a consistently greater mean duration (140 vs. 110 ms), a higher center of spectral gravity (2823 vs. 1499 Hz), as well as a higher standard deviation of the center of spectral gravity (4208 vs. 2457 Hz), suggesting a much wider distribution of noise energy. On the other hand, the differences in the intensity, skewness and kurtosis were not consistent between /ð/ and /v/ and no clear mean cross-context effects were found for either consonant.

More pertinent to the vocalic context specific pattern of /ð-v/ confusions, in the /u/ context, the differences between /ð/ and /v/ spectral center of gravity were considerably larger, particularly in the second half of the noise signal, than in the other two contexts (i.e., 2407 vs. 4716 Hz, for the /ð/ and /v/ respectively, in the /u/ context as opposed to 682 vs. 1025 Hz for /ɑ/ and 490 vs. 2370 Hz for /i/). The higher spectrum and greater spectral differences in the /u/ context could render it easier for the listeners to identify the correct place of articulation for each consonant. However, interspeaker variability on these measures remained high, suggesting that listeners might be adaptively using a variety of acoustic cues to detect appropriate articulatory gestures (Flege et al., 1997).

It also cannot be ruled out that the superior /ð/ performance in the /u/ context was influenced in some ways by the fixed order in which vocalic contexts were presented to the listeners (consistently as /ɑ/, followed by /i/, followed by /u/). However, if practice contributed to performance accuracy, its effects would be expected for the /ɑ/ and the /i/ contexts as well. The absence of improvement in perceptual accuracy of /ð/ between the first (/ɑ/) and the second (/i/) vocalic context for either group (the /ð/ accuracy actually decreased by 24 and 27 points for NA and NE groups, respectively), suggests that an improvement in accuracy in the third /u/ context by 42 and 59 points would be rather unlikely. Furthermore, positive effects of presentation order would be expected to apply to other consonants, as well, and cannot explain the number of /r-v/ errors that appear only in the /u/ context for both groups. Thus, presentation order effects, while conceivable, would not be likely to have a strong influence on the results. It should also be noted that the stimuli in Shannon et al. (1999) did not include the voiceless interdental fricative. Future studies should include /θ-f/ comparisons in order to determine whether the confusions surrounding this voiceless pair mimic their voiced counterpart in varying as a function of vocalic context.

4.4 Limitations and future directions

Native Arabic bilinguals in this study demonstrated an overall high accuracy level in the perception of AE speech sounds, despite considerable perceptual difficulties with some specific vowels and consonants. However, the pattern of errors does not appear to be clearly linked to native Arabic phonology (e.g., there were no /p-b/ confusions despite the absence of /p/ in Arabic). The similarity in overall performance and specific identification errors of the listeners in NA and NE groups tentatively suggests a greater role of the L2 local dialects in the perceptual classification of AE sounds. Nevertheless, this conclusion needs to be examined further with a group of less skilled English listeners whose native language is Arabic, who would likely produce a greater number of errors than the listeners in the present study, thus allowing finer differentiation of the effects of native phonology between groups. Indeed, participants’ high identification accuracy might have obscured some of the more subtle differences in perception of vowels and consonants that could be based on the differences in the listeners’ native phonologies.

Perhaps, given sufficient time in the UAE, residents perceive sounds in a consistent manner. That is, over time and with frequent interaction, speakers of different dialects who speak English as a lingua franca may converge on a single, more phonologically stable, dialect of English (Orr et al., 2011). Although this study included participants with heterogeneous Arabic and English
dialect backgrounds, most of the listeners had resided in the UAE for several years. Unfortunately, no comprehensive description of the phonological characteristics of English dialects spoken in the UAE or neighboring countries exists. The present study points to a need for such descriptions and highlights the complexity of the question of whether dialectal differences can be separated from native language interference in a setting where diverse English and Arabic dialects interact on a daily basis. It is hoped that future research will provide systematic descriptions of local UAE English dialects, examine the effects of specific vowel and consonant confusion patterns, and detail perceptual relationships among vowels and consonant systems specific to individual English and Arabic dialects (e.g., see Clopper, Levi, & Pisoni, 2006, re: perceptual similarity across AE dialects). Such studies would enhance our understanding of changes in phonological representation that occur in the course of learning a second language.

Age of acquisition has been shown to be a powerful predictor of ultimate perceptual mastery of second language speech sounds (e.g., Flege, MacKay, & Meador, 1999). Thus, one reason for the lack of clear native Arabic language interference effects is likely to be the young age at which the native Arabic participants started to learn English. Future research might examine comparisons between early and late Arabic-English bilinguals in the region, matched, to the extent possible, on socio-economic characteristics and dialectal background, in order to investigate more precisely the effects of native Arabic language interference on the perception of AE speech sounds.

Another possible confound is that the key words representing the vowels differed in their lexical frequency, with ‘had’, for example, being far more frequently said than ‘hud’. However, these highly proficient listeners would be expected to be very familiar with English sound-spelling correspondence and would thus be expected to select the responses based on the corresponding written vowel in the key word rather than on the lexical frequency of the key word.

4.5 Implications for communication

An understanding of the potential interference effects, or lack thereof, of Arabic as the first language on the perception and production of English can be utilized in teaching English as a second language to native speakers of Arabic. Syed (2003) addresses the need for research-based knowledge to develop informed pedagogy as one of the main four central contributing aspects to enhance teaching English as a second language in the Gulf region. The present study suggests that it may be prudent to pay attention to factors such as vocalic context, for example, while working with a native Arabic-speaking student showing difficulty in perceiving the /ð/ in English.

The finding of particular difficulty with AE vowel perception in a country where AE is spoken frequently points to the importance of being able to adjust to dialectal variations in order for communication to be successful. Clearly, being a proficient early learner of English or a speaker of a local English dialect does not prevent listeners from having difficulties in the perception of speech sounds from a different dialect. These perceptual difficulties, and associated communication problems, may be further exacerbated by suboptimal communication settings such as reverberant and noisy lecture halls or by hearing impairment, which would further limit or distort perceptually salient acoustic cues. Similarly, the accuracy of clinical assessment of speech perception abilities (i.e., speech audiometry) may also be affected by native language and native dialect-specific speech sound confusions (Calandruccio, 2010). Work on training listeners on perception of dialectal variations (Clopper & Pisoni, 2004) and perceptual adaptation to accent (i.e., Bradlow & Bent, 2008) may enhance the comprehension of several varieties of a language in regions in which that language is required for communication.
Acknowledgements

The authors wish to thank Lauren Liria, J.D. Purdy, Cristina Sanchez, Valerie Levy, Elanna Seid, and Wid Al Bayaty for their helpful suggestions and assistance with preparation of the manuscript. We are grateful to Drs. Shannon and Hillenbrand for making their stimuli available for this study and to the participants in this study for their kind cooperation.

Note

1. To determine whether additional proficiency in Arabic by four of the participants in the NE group could have affected the identification results, all error analyses described in the Results section below were repeated without these participants. However, no significant differences in identification error patterns emerged: for any vowel or consonant sound the difference in scores between all NE participants and those without NE participants who also spoke Arabic was always smaller than 5 percentage points. Therefore the NE participants who spoke Arabic were kept as part of the NE group.

References


