Full Research Article

LANGUAGE TEACHING RESEARCH

Language Teaching Research
1–27

DOI: 10.1177/1362168820931888
journals.sagepub.com/home/ltr
SAGE

The English pronunciation of Arabic speakers: A data-driven approach to segmental error identification

Ivana Rehman
Alif Silpachai
John Levis
Iowa State University, USA

Guanlong Zhao
Ricardo Gutierrez-Osuna
Texas A&M University, USA

Abstract

The accurate identification of likely segmental pronunciation errors produced by nonnative speakers of English is a longstanding goal in pronunciation teaching. Most lists of pronunciation errors for speakers of a particular first language (L1) are based on the experience of expert linguists or teachers of English as a second language (ESL) and English as a foreign language (EFL). Such lists are useful, but they are also subject to blind spots for less noticeable errors while suggesting that other more noticeable errors are more important. This exploratory study tested whether using a database of read sentences would reveal recurrent errors that had been overlooked by expert opinions. We did a systematic error analysis of advanced L1 Arabic learners of English (n = 4) using L2 Arctic, a publicly available collection of 1,132 phonetically-balanced English sentences read aloud by 24 speakers of six language backgrounds. To test whether the database was useful for pronunciation error identification, we analysed Arabic speakers' sentence readings (n = 599), which were annotated in Praat for pronunciation deviations from General American English. The findings give an empirically supported description of persistent pronunciation errors for Arabic learners of English. Although necessarily limited in scope, the study demonstrates how similar datasets can be used regardless of the L1 being investigated. The discussion of errors in pronunciation in terms of their functional loads (Brown, 1988) suggests which persistent errors

Corresponding author:
Ivana Rehman, Iowa State University, Ross Hall, 527 Farm House Lane, Ames, IA 50014, USA
Email: ilucic@iastate.edu

are likely to be important for classroom attention, helping teachers focus their limited classroom time for optimal learning.

Keywords
Arabic learners, error frequency, functional load, substitution, deletion, insertion, pronunciation corpus, pronunciation teaching, segmental errors, teaching priorities

I Introduction

A perennial concern in pronunciation teaching is the accurate identification of likely production errors by second language (L2) speakers. Identification can best be done learner by learner, but such an approach is time-consuming and assumes no commonality between learners. Other approaches to identifying errors have focused on learners' first language (L1) backgrounds, with the assumption that mismatches between the phonological systems of the L1 and target language (TL) will help identify common difficulties. This was called the contrastive analysis hypothesis (CAH; see Munro, 2018). Although the effect of L1 is noticeable for all linguistic features, it is often thought to be 'most striking in the case of pronunciation . . . giving rise to what we call, for example, a Dutch or Turkish, or Chinese "accent" ' (Swan and Smith, 2001, p. xi). The CAH, however, was not precise. Some predicted errors did not occur while other errors not predicted by the contrastive analysis did occur, often across L1s, as part of the developmental path shared by learners (Munro, 2018). In addition, some language learners successfully produce unusual or difficult L2 sounds without trouble depending on how different they are from sounds in the L1 (Best & Tyler, 2007). In other words, while L1 background could, predict difficulties to some extent, it could not predict all errors (Wardhaugh, 1970) of pronunciation (Celce-Murcia et al., 2010; Munro, 2018).

In addition to the CAH, other approaches have been used for pronunciation error identification, especially expert analysis. Nilsen and Nilsen (1971), for example, used 'more than fifty linguists and other language specialists' (p. xiii) to identify likely pronunciation errors according to the L1 of the learners. Similarly, Swan and Smith (2001) drew on the expertise of various writers in describing the grammatical and phonological errors associated with English learners from various L1s. Authors have also written about errors for speakers of particular L1s such as Arabic and Vietnamese. In such books and articles, the role of expertise is central: 'Guided by our experience in teaching English to Arab students . . . we have attempted to identify the problems . . . learners continue to make' (Kharma & Hajjaj, 1989, p. 2). Expertise may or may not draw upon carefully collected data. For example, one expert analysis occurred 'not only in collecting types of problems and mistakes from hundreds of papers, compositions, summaries and translation . . . over an extended period of time [but] we cannot claim to be fully comprehensive. There might be a few minor problems which have not been tackled at all' (Kharma and Hajjaj, 1989, p. 3). Interestingly, this quotation makes no mention of pronunciation errors, but the book nonetheless includes a chapter about them.

However, most lists of pronunciation errors for speakers of particular L1s are based on anecdotal evidence and the experience of language teachers but do not include data supporting the classifications of errors (see McAndrews & Thomson, 2017). In one book, the author admits that ‘it is impracticable to attempt in a book of this kind to cover all the errors of pronunciation that are commonly made by Arabs as individuals and as representative of their own many different regions and countries’ (Mitchell and El-Hassan, 1989: 11). For example, Avery and Ehrlich (1992) list segmental and suprasegmental errors that are likely for many different L1s, along with brief reasons why they are problematic, e.g. ‘Arabic does not have a contrastive /v/ sound, although the sound does occur as a positional variant of /f/ before voiced stops and fricatives’ (p. 111). But like other lists of errors, they include no citations or data about how the lists were compiled.

Even if lists of errors are accurate, they often imply that all errors are equally valuable as teaching targets, missing a key aspect of a principled approach to setting priorities for teaching (McAndrews & Thomson, 2017). Korean speakers of English often have trouble with both the /d–ð/ distinction and the /l–ɹ/ distinction. While both deviations are likely to mark foreign accentedness, the /l–ɹ/ confusion is more critical to intelligibility (actual understanding of a speaker; see Levis, 2018), that is, it is far more likely to result in confusion for listeners. The /d–ð/ distinction, in contrast, is unlikely to result in loss of intelligibility, serving to mark a deviation from native norms instead (Levis, 2005).

II Literature review

1 Segmental pronunciation and error gravity

The seriousness of segmental pronunciation errors (deviations from a norm based on a particular variety of speech) has been determined in various ways. Prator and Robinett (1985) write that the cumulative frequency of phonetic deviations results in loss of intelligibility. Unfortunately, they provide very little information about what levels of errors should be considered frequent nor about what types of deviations they considered. However, a more precise predictor of error gravity within a language is functional load (FL) (see Brown, 1988; Catford, 1988; Gilner & Morales, 2010; Munro & Derwing, 2006; Surendran & Niyogi, 2003, 2006). One crucial distinction between frequency and FL is that the latter encompasses semantic contrasts. That is, functional load only makes sense in reference to particular phonemic contrasts. At a basic level, FL measures the number of minimal pair contrasts distinguished in a language. In Brown (1988), FL is measured on a scale of 1–10, with 10 being high. Thus, high FL contrasts occur more than low FL contrasts. In English, for example, there are more minimal pairs for /p/–/b/ (e.g. *pat~bat*) than for /ʃ/–/ʒ/ (e.g. *mesher~measure*) (Hockett, 1966).

For L2 pronunciation, FL is important to teaching. Catford (1987) argued that FL should be used to guide an optimal selection of segments to be taught. Munro and Derwing (2006) suggested that the teaching of high FL contrasts may be particularly important for comprehensibility (i.e. the amount of work a listener must do to understand a speaker in a given context; see Levis, 2018). In Munro and Derwing (2006), judgments of comprehensibility (the amount of work listeners have to do to understand) were much worse for high FL errors than for low FL errors. In regard to accentedness, sentences with errors in lower FL contrasts (e.g. /d/–/ð/) evoked similar judgments of accentedness for one, two or three errors. In contrast, judgments of accentedness for sentences with errors in high FL contrasts (e.g. /l/–/n/) increased for multiple errors. Although accentedness is not central to our study, the changes in accentedness judgments for multiple errors in individual sentences may indicate that more frequent errors in connected speech may affect a speaker's comprehensibility. Indeed, Munro and Derwing's (2006) results show a nonsignificant trend toward this.

Other approaches to error gravity have looked at the communicative breakdowns caused by pronunciation errors. Jenkins (2000) found that pronunciation errors were implicated in a majority of communication breakdowns during English interactions between nonnative speakers. Most of these breakdowns involved errors in segmentals, a finding echoed by Deterding (2013) in interactions in English by speakers from Southeast Asia. The impact of segmental errors on intelligibility, especially those in stressed syllables, has also been established by Zielinski (2006), Im and Levis (2015), and Uzun (2019). In these studies, listeners were asked to listen to speech produced by L2 learners and identify where they struggled to understand and why.

Not all segmental errors involve phonemic contrasts, however, and functional load tells us little about how listeners are affected by these errors. For example, an Arabic error consistently mentioned by experts (/r/ for /ɹ/) is unlikely to be mistaken for another phonemic category but is highly noticeable to listeners. Other segmental errors that cannot be evaluated according to functional load include those that affect syllable structure, especially those in which segments are deleted or where segments are inserted to make consonant clusters easier to pronounce. There is uncertainty about the impact of these insertion and deletion errors on intelligibility. Anderson-Hsieh et al. (1992) found that changes in syllable structure due to deletion and insertion resulted in reduced pronunciation ratings. Epenthesis errors in English, which may occur for Arabic speakers because of markedness constraints (Alezetes, 2007), can be seen as a type of segmental error that affects intelligibility. Epenthesis errors that affect understanding have also been argued for Korean learners of English (No, 1997) and Cantonese learners of English (Chan & Li, 2000). Deletion errors also affect syllable structure. Vietnamese speakers have been documented deleting segments in final consonant clusters in English words (Nguyen, 2008), a strategy that strongly impaired intelligibility (Zielinski, 2006).

2 Arabic pronunciation of English

Arabic is a native language of over 300 million people across thousands of miles in approximately 25 countries. It is also used as a second language by millions more. Arabic is diglossic, with Modern Standard Arabic, based on Classical Arabic, used as the written and formal spoken standard alongside multiple dialects of spoken Arabic that differ from the written standard in varying ways (https://en.wikipedia.org/wiki/Arabic).

Like other geographically distributed languages (e.g. Chinese, English), not all of these spoken varieties are mutually intelligible (Čéplö et al., 2016). While it is likely that Arabic speakers from different spoken varieties may face different challenges in pronouncing English, expert views on Arabic speakers' pronunciation of English have identified a number of likely segmental difficulties that impact all speakers of Arabic.

A number of books (Al-Mutawa & Kailani, 1989; Avery & Ehrlich, 1992; Kharma & Hajjaj, 1989; Mitchell & El-Hassan, 1989; Swan & Smith, 2001) have identified varied problems including consonant errors such as /p/ pronounced as /b/, /v/ as /f/, deviations for final /ŋ/, distinguishing the vowels in *bit–bet*, and insertions of vowels in initial consonant clusters. These experts also differ in the types of errors they identify. Tables 1–3 list consonant, vowel and syllable structure errors previously identified for Arabic learners of English. Due to different ways of talking about pronunciation errors, different authors classify errors differently, such as identifying difficulties with particular contrasts (e.g. /p/–/b/) and identifying particular classes of sounds as problematic (e.g. voiced–voiceless distinctions), but without suggesting whether such problems are likely to occur for all such distinctions in English. In addition to specific consonant and vowel errors, the experts sometimes suggest potential causes for errors, including challenges with producing fortis–lenis distinctions (Mitchell & El-Hassan, 1989), or using dental rather than alveolar places of articulation (Kharma & Hajjaj, 1989; Mitchell & El-Hassan, 1989; Mutawa & Kailani, 1989).

Phonotactic rules exclusive to English have also been suggested to be a source of Arabic speakers' pronunciation errors. Learners often insert a vowel, which alters the syllable structure, when they have not mastered English phonotactics. This can be seen in certain consonant clusters, e.g. /spɹ, skɹ, stɹ, pɹ, gɹ/, etc. which are broken up by vowel insertion, e.g. *spring*s pronounced *ispring* or *sipring* (Avery & Ehrlich, 1992; Kharma & Hajjaj, 1989; Smith, 2001). Learners may also insert a consonant, making the syllable conform to Arabic phonotactics. For example, /ŋ/ is often pronounced as /ŋg/ or /ŋk/ (Avery & Ehrlich, 1992; Kharma & Hajjaj, 1989; Smith, 2001).

Lastly, the influence of English spelling on pronunciation has also been discussed. Arabic-speaking learners reportedly read words phonetically, given that the Arabic writing system is sometimes argued to be so regular in its sound–spelling correspondence (Kharma & Hajjaj, 1989), that it has been called 'virtually phonetic' (Smith, 2001, p. 198), despite its vowel sounds not always being represented in writing. As a result, English words may also be pronounced as though English is equally transparent in its orthography. For example, *walked* may be read as /wɔːkɪd/ (Kharma & Hajjaj, 1989, p. 19), or words may be subject to consonant doubling, e.g. /əˈlaʊ/ *allow* is pronounced as /ælˈlaʊ/ (Kharma & Hajjaj, 1989, p. 12).

Table 2 shows the vowel errors listed as likely for Arabic speakers of English. Only one contrast, the vowels in *bit–bet*, was named by all experts, five of 11 were named by three or more, and the rest were put forth by two or fewer. As with consonant errors, some errors are salient enough to be noticed by multiple experts, while other seem less noticeable even to experts.

A third set of expert-identified errors involves the pronunciation of sounds in relation to syllable structure (Table 3). In all cases, these involve insertion, the insertion of vowels within consonant clusters, either in general or clusters created by the use of past tense morphemes, or the insertion of consonants in words with /ŋ/, where *thing* may sound like *think* or *singer* like *singger*. Another error classified as insertion was when Arabic speakers may pronounce both spelled consonants (e.g. the <pp> in *apple*) separately.

3 Research questions

The purpose of this study is to examine the usefulness of undertaking a systematic analysis of pronunciation errors using a phonetically-annotated corpus/database. We hypothesize that even a modestly-sized database of speech can offer a new view on the distribution

Table 1. English consonant errors for speakers of Arabic in various sources.

Error (FL from Brown, 1988)	Kharma & Hajjaj (1989)	Mitchell & El-Hassan (1989)	Al-Mutawa & Kailani (1989)	Avery & Ehrlich (1992)	Swan & Smith (2001)
Consonant contrasts:					
/f/–/v/ (7)	×	×	×	×	×
/tʃ–ʃ/ (2)	×	×		×	×
/dʒ–ʒ/ (n/a)	×	×		×	×
/p/–/b/ (10)	×	×	×	×	×
/k/–/g/ (9)		× (Cairo Arabic)			×
/s or t–θ/ (5, 4)	×		×	×	×
/z or d–ð/ (7, 5)	×		×	×	×
Other consonant errors:					
Wrong variants of /r/	×	×	×	×	×
Overuse of dark /l/	×	×			
Syllabic consonants	×	×			
Lack of aspiration	×	×			
Nasal plosion		× (e.g. *hypnosis*)			
Lateral plosion		× (e.g. *badly*)			

Notes. FL = functional load. × means that the error is discussed by the authors. Numbers in parentheses following contrasts indicate the functional load of the contrast using Brown (1988). The numbers range from 1 (lowest) to 10 (highest).

Table 2. English vowel errors for speakers of Arabic.

Error (FL from Brown, 1988)	Kharma & Hajjaj (1989)	Mitchell & El-Hassan (1989)	Al-Mutawa & Kailani (1989)	Avery & Ehrlich (1992)	Swan & Smith (2001)
Vowels in *bit–bet* (9)	×	×	×	×	×
Monophthongization of diphthongs	×	×	×		×
Vowels in *luck–lock* (10)	×	×		×	
Vowels in *beat–bit* (8)		×	×	×	
Vowels in *fool–foot* (3)	×	×		×	
Vowels in *dirt–dart* (n/a)	×	×			
Vowels in *sale–sell* (9)	×			×	
Vowels in *boat–bought* (10)	×	×			
Pronunciation of schwa	×	×			
Vowel in *bat (bet)* (10)		×		×	
Vowels in *cot–caught* (4)					×

Note. FL = functional load.

Table 3. English syllable structure errors for speakers of Arabic.

Error*	Kharma & Hajjaj (1989)	Mitchell & El-Hassan (1989)	Al-Mutawa & Kailani (1989)	Avery & Ehrlich (1992)	Swan & Smith (2001)
Consonant cluster insertion	×	×	×	×	×
Past tense vowel insertion (e.g. *walked*)	×		×		×
/ŋ/ → /ŋg/ or /ŋk/	×	×	×	×	×
Pronouncing geminate consonants (e.g. *ap-ple*)	×		×		

Note. * No functional load (FL) for these types of errors.

of pronunciation errors for speakers of a particular L1. Positive results would provide a justification for including more widely representative phonetic corpora stratified by spoken proficiency and by task type. Specifically, this study of Arabic pronunciation errors uses the L2 Arctic corpus based on phonetic annotation of 599 sentences from four advanced Arabic speakers of English to answer two research questions.

1. How frequent are errors of substitutions, distortion, deletions, and insertions?
2. To what extent do errors found in the data-driven approach differ from the errors identified by experts? How do differences reflect upon the strengths and weaknesses of the two approaches?

III Methodology

We hypothesize that using a consistent task, considering the sampling of speakers and their characteristics, and phonetically annotating errors will reveal patterns of errors that differ from expert views. It is ideal to use a phonetically-annotated learner corpus to carry out this analysis (e.g. Van Doremalen, Cucchiarini & Strik, 2009), although these are exceedingly rare (Biber, 2019), not easily available (O'Brien et al., 2018; Zhao et al., 2018) or include broad phonetic transcriptions that are not easily searchable (O'Brien et al., 2018). Although it would be valuable to have a carefully constructed corpus of authentic speech from a specified discourse domain, the use of L2 Arctic offers a way to test what such a corpus could reveal about learners' pronunciation errors.

In this study, we employed a phonetically-annotated collection of read aloud sentences from L2 Arctic. Although there are reasons to argue that L2 Arctic is not a corpus, Zhao et al. (2018) describe it as one. We will likewise call it a corpus and discuss issues related to corpus definitions at the end of the article. L2 Arctic is a publicly available database of 1,132 English sentences read aloud by 24 speakers: four speakers (two male, two female) from six different language backgrounds (Arabic, Chinese, Hindi-related, Korean, Spanish; https://psi.engr.tamu.edu/L2-Arctic-corpus). The 1,132 sentences were the same as those from the CMU ARCTIC set (http://festvox.org/cmu_arctic). According to the CMU Arctic website, the set of sentences, chosen from Project Gutenberg texts for which copyright had lapsed, was designed to be phonetically balanced. This CMU Arctic sentences, originally recorded by native speakers of English for use in speech synthesis applications, were re-recorded by nonnative speakers for L2 Arctic because they were likely to include multiple examples of all English sounds and contexts and would be challenging for a wide range of nonnative English speakers, making them suitable for research on areas such as automatic mispronunciation detection or speech synthesis. In this article, we used them to examine the distribution of pronunciation errors for a single L1. All speakers were advanced speakers of English (most were studying in either undergraduate or graduate programs at an American Research I university) who could read aloud the sentences during one recording session. The purpose of this study is to help understand whether even a modest phonetically-annotated sample can help identify common L2 segmental pronunciation errors for English speakers from one L1, Arabic.

The sample used in this study consists of 599 sentences (150 sentences read by two male and two female advanced Arabic speakers of English, less one sentence not read by

Table 4. Demographic information for Arabic participants from second language (L2) Arctic.

	ABA	ZHAA	SKA	YBAA
Age (years)	32	40	26	24
Sex	Male	Female	Female	Male
Country of origin	Egypt	Saudi Arabia	Egypt	Jordan
Length of residence	1 year	6 years	1 year	2 months
Age of onset (years)	10	14	10	6

one speaker). The sentences were phonetically annotated in Praat for pronunciation deviations from General American English. The resulting lists of errors are then discussed in light of their frequency and functional load (Brown, 1988). Demographic information for the Arabic speakers is in Table 4.

Although L2 Arctic includes recordings of all 1,132 sentences for each speaker, only 150 sentences were phonetically annotated for each speaker because of the work involved in aligning the texts with the recordings and annotating the textgrids. According to the L2 Arctic documentation, the first 100 sentences in the database were the same for speakers from all L1 backgrounds to promote comparisons across L1s while the next 50 sentences were included to elicit likely L1-specific errors. The recordings from all 150 sentences per speaker were aligned to the written text using the Montreal forced-aligner, a program that creates a time-aligned version of a written text and an audio file (https://montreal-forced-aligner.readthedocs.io/en/latest).

The forced-aligner was used to create Praat textgrid transcriptions for each sentence containing word and phone boundaries. Phones were written in ARPAbet to facilitate computer processing (Appendix 1). Each TextGrid file was manually checked and annotated by three phonetically-trained annotators who underwent training and calibration with each other and a project supervisor. For each sentence, annotators corrected all boundary and/or ARPA phonetic transcription errors. In this article, we use both ARPA symbols and IPA symbols. In a comment tier of the TextGrid, IPA symbols were used to portray the errors. The annotators classified deviations from the expected sounds in four ways: as substitutions, insertions, deletions, and distortions. Substitutions were an expected phoneme being classified as another English phoneme, insertions were the addition of a new sound to the expected sound, deletions the non-pronunciation of an expected sound, and distortions the pronunciation of an expected English phoneme with an identifiable non-phonemic variant. In cases where annotators were not certain, the transcription was discussed with the other annotators and the project supervisor.

IV Results

The 599 sentences for the four Arabic speakers included 19,764 individual phones (4,941 phones/speaker, or an average of 33 phones per sentence). There were 2,202 segmental errors (1,305 substitutions, 403 distortions, 220 deletions, and 274 insertions), totaling 11.14% of all segments, or an average of 3.68 errors per sentence. The first research question asked about the error distribution for substitutions, distortions,

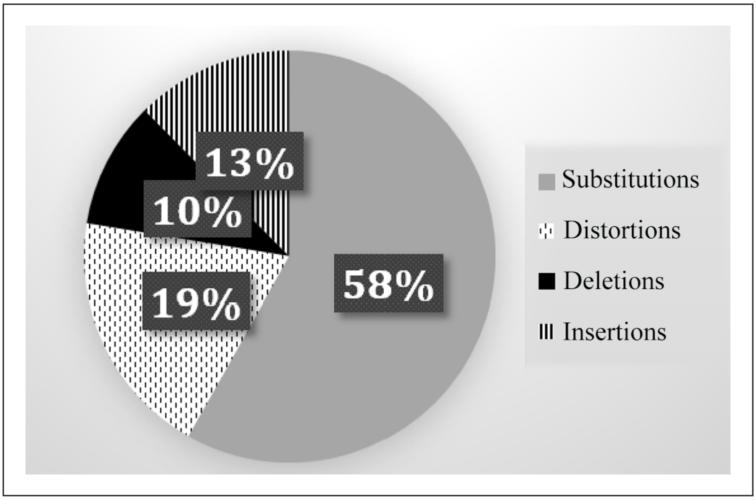

Figure 1. Phoneme error distribution.

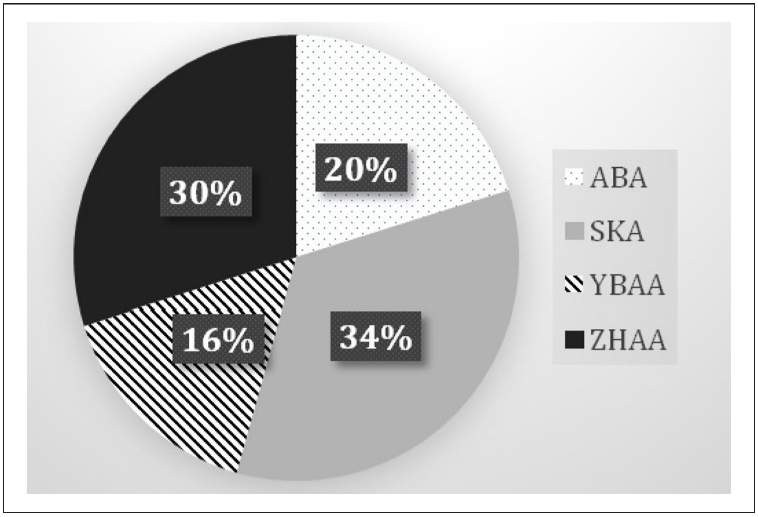

Figure 2. Substitution error distribution.

deletions, and insertions. As shown in Figure 1, substitutions comprised well over half of all deviations. Distortions, which in this corpus were primarily the use of /r/ or /ɾ/ for /ɹ/, were not counted as a substitution because the Arabic <r> sounds are not phonemic in English. Figures 2, 3, 4, and 5 show the distribution of errors across individuals per type of error. Figure 2 shows the distribution of substitution errors per speaker: SKA produced 34%, ZHAA produced 30%, ABA produced 20% and YBAA produced 16% of all substitution errors.

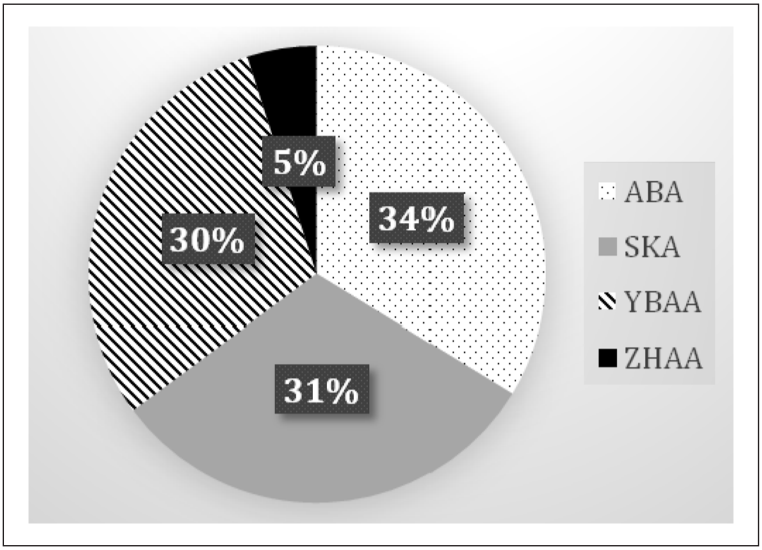

Figure 3. Distortion error distribution.

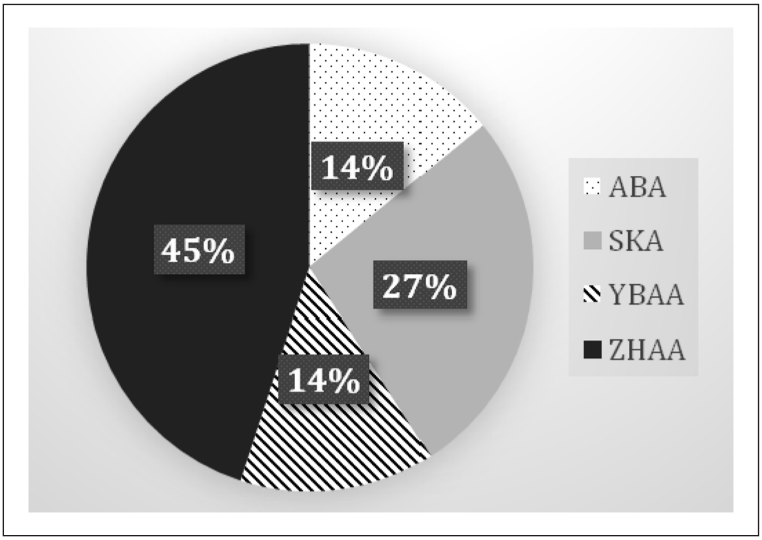

Figure 4. Deletion error distribution.

Distortion errors are presented in Figure 3, and show a fairly equivalent distribution for three out of four participants. ABA, SKA, and YBAA produced 34%, 31% and 30% of all distortion errors, respectively, largely because ZHAA was more successful in her production of /ɹ/ resulting in a total of 5% of distortion errors.

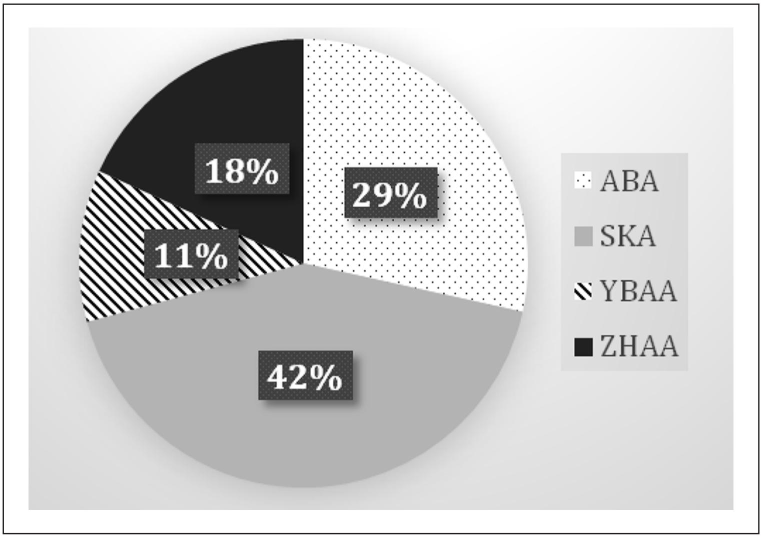

Figure 5. Insertion error distribution.

Deletion errors are presented in Figure 4. ZHAA, who had the lowest proportion of distortion errors, had the highest number of deletion errors (45%), while other participants had fewer: SKA with 27%, and ABA and YBAA with 14% each.

Insertion errors (Figure 5) follow a similar distribution pattern as deletion errors, with two speakers producing more and two less. SKA produced 42% of all insertion errors. ABA produced 29%, followed by ZHAA and YBAA, with 18% and 11%, respectively.

1 Substitutions and distortions

Substitutions and distortions, in which the expected production was interpreted as another identifiable sound, were the most common type of pronunciation error. Some of the sounds were more prone to deviate from what is expected than others, but there were a number of errors that were common enough to suggest that they are likely to persist in the speech of advanced proficiency learners. By percentage, the most common phoneme error, /dʒ/ or JH in ARPAbet symbols, was substituted 69 times out of a total of 172 occurrences (40.1%). This phoneme was mostly substituted by /ʒ/ (47 times, or 68.1%), or /tʃ/ (21 times, or 30.4%); /ʒ/, even though it had a 25% error rate, only had a total of 20 occurrences.

The phoneme /ɹ/ had 314 errors (35.1%) out of 894 occurrences. We classified this as a distortion because the participants mostly did not substitute /ɹ/ with a different phonetic category. In Arabic, <r> is an alveolar trill, so L1 Arabic learners of English tended to produce it in English instead of the postalveolar approximant. In our dataset, the alveolar trill [r] was used instead of the postalveolar approximant /ɹ/ 289 times out of a total of 314 substitutions (92%). Additionally, our dataset shows 113 mispronunciations of [ɚ] out of 496 occurrences (22.8%), and most of these pronunciations were due to participants' use of [r] rather than a postalveolar approximant or the rhotacized vowel.

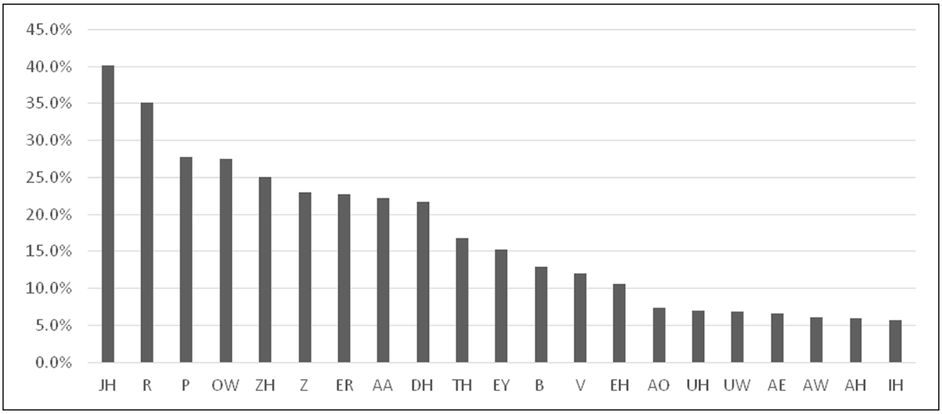

Figure 6. Distribution of substitution errors
Note. Phonemes with less than 5% substitution percentage distribution are excluded from this graph.

The phoneme /p/ had a total of 392 occurrences and 109 substitutions (27.8%), with /b/ as 101 of these substitutions (92.7%). As /p/ is not phonemic in Arabic, L1 Arabic learners of English tended to substitute the closest phoneme that exists in their phonetic inventory, which is /b/. Additionally, even though /b/ was substituted 44 times out of 339 occurrences (13%), 100% of those substitutions were /p/, suggesting that /p/–/b/ is a bidirectional error for advanced learners.

Monophthongization occurred with the phoneme /oʊ/. Out of 313 total occurrences, 86 were substituted (27.5%), and 72 (83.7%) of these substitutions were /ɔ/. Another incidence of monophthongization is /eɪ/, which had 57 errors of 375 occurrences (15.2%). Of these 57 substitutions, 56 involved a monophthong, mostly /ɛ/ (44 times, or 77.2%); /ɛ/ was also substituted for /ɪ/ (23 times) or /i/ (12 times).

The vowel /ɑ/ occurred 243 times, and it was substituted 54 times (22.2%). Twenty-two (40.7%) of these substitutions were with /ɔ/, which is an acceptable variant in American English. However, 18 (33.3%) of these substitutions were with /oʊ/, and this mispronunciation may be characteristic of spelling pronunciation, as /ɑ/ is often spelled with <o> in English. Errors with other vowels were infrequent, with less than 10% of the tokens substituted by another segment.

The fricative /ð/ occurred 594 times, and there were 129 substitutions (21.7%). These substitutions were mostly /z/ (63 substitutions, or 48.8%) and /d/ (55 substitutions, or 42.6%); /θ/ had 23 substitutions of a total of 137 occurrences (16.8%), and most (78.3%) of these substitutions involved /s/.

An interesting observation in our dataset is that, even though /v/ occurred only 46 times, it was always pronounced as /f/. Previous literature lists /f/–/v/ as a bidirectional pronunciation issue for native speakers of Arabic, but our data shows that it was unidirectional because /f/ had only three substitutions out of 440 occurrences (0.7%). The rest of the phonemes were at 5.4% substitution or lower. Figure 6 shows the distribution of substitution errors across phonemes. Figure 7 shows the distribution of substitution and distortion errors across individuals.

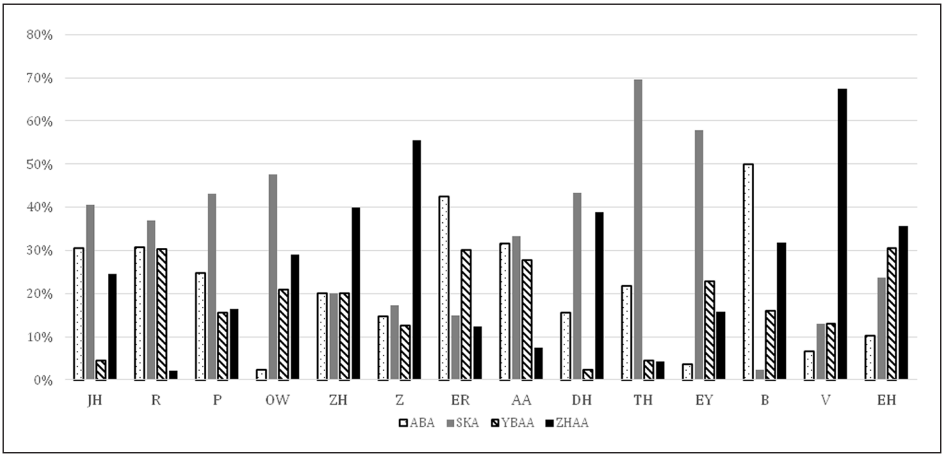

Figure 7. Distribution of substitution errors per individual
Note. Phonemes with less than 10% substitution percentage distribution are excluded from this graph.

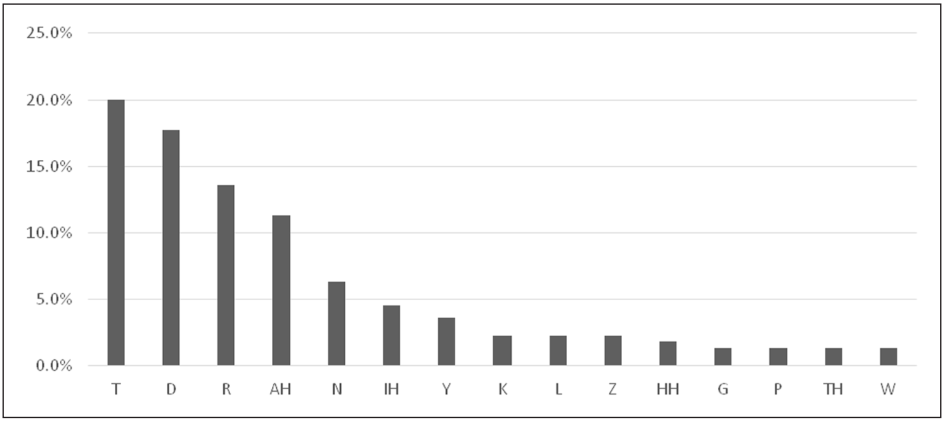

Figure 8. Distribution of deletion errors
Note. Phonemes with less than 1% deletion percentage are excluded from this graph.

2 Deletions

Figure 8 shows the distribution of phoneme deletion errors. Common deletions seemed to involve /d/ and /t/, with 39 deletions out of 907 occurrences (4.3%), and 44 deletions out of 1,183 occurrences (3.7%), respectively. Most of these deletions were related to either the grammatical morpheme which marks past tense (*-ed*), or they occurred at the end of the word when the following word began with dental fricatives or alveolar stops; /ɹ/ was deleted 30 times out of 894 occurrences (3.4%), mainly postvocalically in word-final positions. Given the acceptability of this pronunciation in major varieties of English,

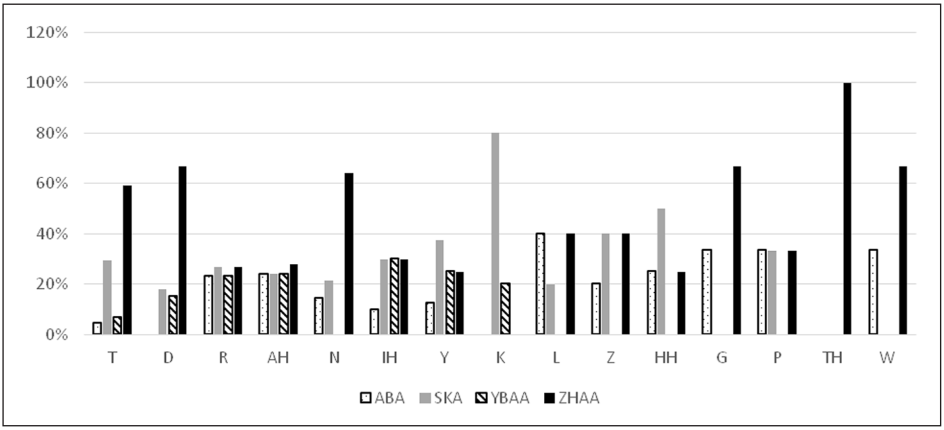

Figure 9. Distribution of deletion errors per individual
Note. Phonemes with less than 1% deletion percentage are excluded from this graph.

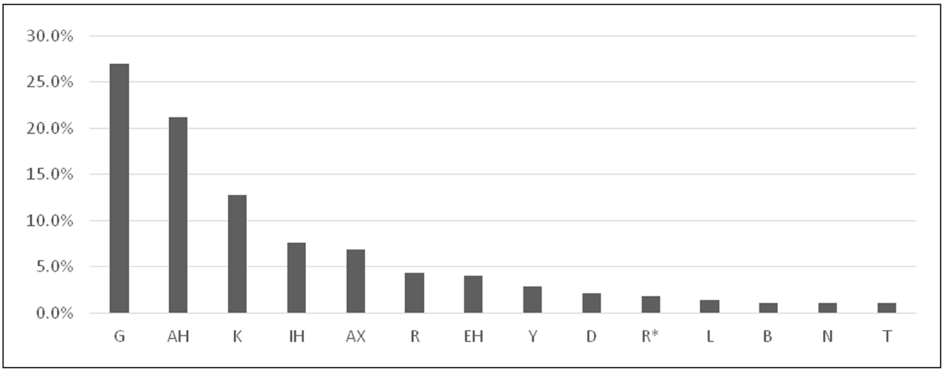

Figure 10. Distribution of insertion errors
Note. Phonemes with less than 1% insertion percentage are excluded from the graph.

such a pronunciation should not be considered an error when looking at pronunciation teaching.

Figure 9 shows the distribution of deletion errors across individuals. As should be expected for advanced learners, each participant had error patterns that marked them as distinct. Some errors were common to all four, while others may have been characteristic of one or two learners only.

3 Insertions

Figure 10 shows the distribution of phoneme insertion errors. There were a total of 274 insertions. We categorized insertions into two groups: insertions involving final /ŋ/ and vowel insertions in consonant clusters. There were a total of 109 insertions of /g/ and /k/

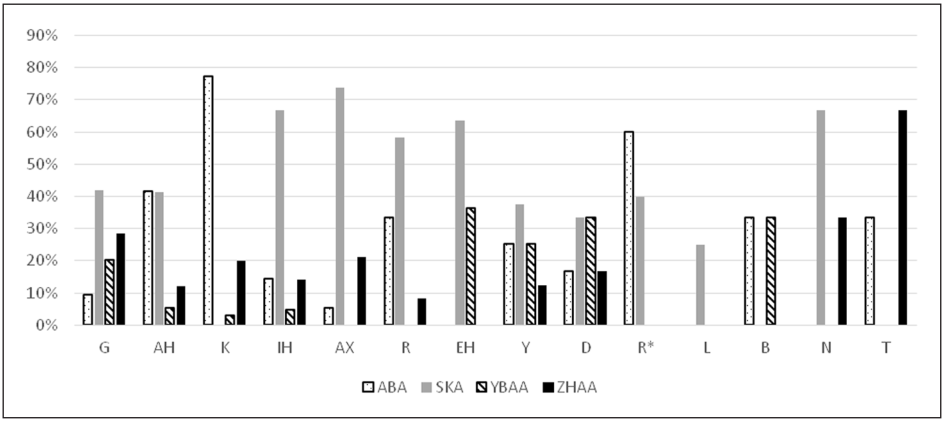

Figure 11. Distribution of insertion errors per individual
Note. Phonemes with less than 1% insertion percentage are excluded from the graph.

sounds (74 and 35, respectively). This amounted to 39.8% of all insertions. All occurrences of [g] and [k] insertions were within words ending in the <ing> suffix. This insertion is likely due to two reasons. First, there is reportedly an orthographic influence on L1 Arabic learners of English, and they tend to read English phonetically. The final <g> is therefore more likely to be pronounced. Second, because the velar nasal is not a phoneme in Arabic, insertions of a homorganic stop make the pronunciation of the final nasal more acceptable for Arabic speakers.

There were also 109 (or 39.8%) insertions of /ʌ/ (58), /ɪ/ (21), /ə/ (19), and /ɛ/ / (11). Out of these 109 occurrences, 102 (or 93.6%) were in consonant clusters. Because English allows more complex consonant clusters than Arabic, these Arabic speakers tended to insert a vowel in between consonants to simplify the pronunciation of the clusters.

Figure 11 shows the distribution of deletion errors across individuals. Similarly to deletion errors, some errors were common to all four, while others may have been characteristic of one or two learners only.

4 Errors identified by the corpus and the experts

The second research question asked about the extent to which errors found in the corpus reflected the errors identified by three or more experts in Tables 1–3. These are reproduced below along with our findings.

Table 5 shows that there was clearly overlap between the two approaches. But the use of L2 Arctic also suggested other error patterns that were not identified by the experts, including mispronunciations of /ɑ/ and /z/. Nor did the experts talk about deletions of sounds identified in L2 Arctic. In addition, three errors identified by multiple experts were uncommon in the L2 Arctic data. This means that, of the 18 errors in Table 5, seven did not overlap. Of the 11 that did overlap, several had differences in the details. For

Table 5. Comparison of expert and corpus findings.

Errors from experts	Our findings	Errors from experts	Our findings
/v/–/f/	[v] only	diphthongs	/eɪ/ → [ɛ], /oʊ/ → [ɔ]
/p/–/b/	[p] → [b], [b] → [p]	*bit–bet*	/ɛ/ → [ɪ], /ɪ/ → [ɛ]
/tʃ/–/ʃ/	This error was unusual	*luck–lock*	An uncommon problem
/dʒ/–/ʒ/	[dʒ] → [ʒ], [dʒ] → [tʃ]	*beat–bit*	An uncommon problem
Variants of <r>	Very common	*fool–foot*	/ʊ/ → [u] (rare)
/s/ or /t/–/θ/	/θ/ → [s]	Not mentioned	/ɑ/
/z/ or /d/–/ð/	/ð/ → [d], /ð/ → [z]	/ŋ/#	[ŋ] → [ŋg], [ŋk]
Not mentioned	/z/ → /s/	Vowel insertion	Common error in consonant clusters
Not mentioned	/ʒ/ → /ʃ/	Not mentioned	Deletions

example, the experts talked about problems with diphthongs, but the only diphthong problems L2 Arctic identified were vowels that are not always classified as diphthongs, /eɪ/ and /oʊ/. No true diphthongs reflected problems. Also, some contrasts reflected errors with only one member of the contrast (/v/), while others showed problems with both sounds in the contrast (e.g. /p/ and /b/). The use of L2 Arctic made these patterns clear whereas the expert judgments were often ambiguous about details of error patterns. Our data also show that some pronunciation features were more characteristic of individual speakers, but the errors above 10% were also problems for most of the speakers in L2 Arctic. It may also be that experts wrote about learners at a lower proficiency level or for other registers of speech, and the infrequency of certain errors in our data would then suggest that they may resolve with increasing spoken proficiency. Overall, the differences suggest that a systematic collection and analysis can more adequately represent the types of errors that occur for a group of L2 learners. A larger phonetically-annotated collection of speech would be able to show whether patterns are different for different proficiency levels or dialects, but the results here suggest that the expert and data-driven approaches often do not overlap.

5 Unexpected errors

The second research question also asked which errors were not identified by experts. There were a number of places in which the corpus revealed other errors, such as the deletion of final [d] and [t], and deletions in environments in which postvocalic /ɹ/ was pronounced as it would be in ɹ-less varieties of English. Some segmental errors that were also found to be more common (above 10% of tokens) included /ɑ/ and /z/. Errors involving /ɑ/ likely included sound–spelling correspondence, as a large number of these included /ɑ/ when it was spelled with <o> or <oa>. A number of the experts indicated that sound–spelling connections are a challenge for Arabic speakers because of the opaque orthography of English (Kharma & Hajjaj, 1989; Mitchell & El-Hassan, 1989; Smith, 2001), and this seems to have been the case with /ɑ/. Some of the /ɑ/ errors involved [ɔ] pronunciations. The Arabic pronunciation of English /o/ was often not a diphthong and was heard as [ɔ]. Since most Canadian and American English speakers do not distinguish [ɑ] and [ɔ], the pronunciation of /o/ as [ɔ] may affect how their words are understood.

Other phonemes identified with errors above 10% of occurrences in the corpus included /z/ and /ʒ/. There are two possible explanations for these errors, which were not directly mentioned by the experts. The first is that Arabic speakers seem to have difficulty producing the voicing distinctions that abound in English. Several experts suggested that Arabic speakers often have trouble with voicing contrasts in English, and this is likely reflected in the speaker's inability to consistently produce voicing contrasts for pairs such as /v/–/f/, /p/–/b/, /k/–/g/, and /z/–/s/. The second issue may have to do with how sibilant distinctions are produced. It is striking that many of the errors discussed among experts and found in our corpus involve distinctions in sibilant sounds (i.e. /z/, /s/, /tʃ/, /ʃ/, /dʒ/, /ʒ/). The patterns for Arabic speakers are different from those documented for Korean speakers of English (e.g. Fox et al., 2009) but, as a class of sounds, sibilants are frequent in English words and carry grammatical meaning for number, possession, and third person singular present verbs at the ends of words. Differences in sibilant inventories or distributions may account for the presence of so many of these sounds in the error counts.

V Discussion

This exploratory study hypothesized that the use of a corpus of read sentences by advanced proficiency Arabic learners of English would provide a useful methodology to identify pronunciation errors, and that it would confirm expert opinions about some errors, disconfirm other opinions, and reveal recurrent errors that had previously been overlooked. Finally, we hoped to use the results to help identify priorities for pronunciation instruction for Arabic learners.

1 Distribution of errors in the corpus

The first research question asked about the error distribution for substitutions and distortions, deletions, and insertions. Unsurprisingly, substitutions and distortions constitute the great majority of segmental errors for L2 learners. Lacking a consistent mastery of particular L2 phonemes or the ability to produce them in particular linguistic environments, L2 learners are likely to produce the sound that is most similar to the target sound in perception and/or articulation. The ubiquity of substitution errors for advanced L2 learners indicates that the inability to consistently distinguish some phonemic contrasts persists beyond the accuracy shown in other contrasts, and that progress in pronunciation learning requires more than increased exposure, especially given that all of these participants were living in an English-dominant environment.

The findings also show that deletion and insertion errors are not uncommon for advanced Arabic learners, and that both types of errors are restricted in the environments in which they are likely to happen. Both insertions and deletions are also likely to affect the ability of a listener to interpret the speech signal accurately by changing the expected syllable structure of a spoken word. It remains an open question whether deletions or insertions are more likely to result in loss of intelligibility, or whether both are important errors. For example, Jenkins (2000) argues that in English as a lingua franca interactions, deletions are likely to harm intelligibility more than insertions since deletions require the reconstruction of segments by the listener. In contrast, insertions include all the expected segments, but they are somewhat masked by the presence of unexpected vowels or consonants. However, it is also possible that some deletions and some insertions impair intelligibility while others do not have the same effect (Levis, 2018). For example, the insertion of a vowel before the s-cluster in ‘spring’ (e.g. *espring*) may be easier for an English-speaking listener to interpret while the insertion of a vowel between the <s> and <p> may be more challenging to interpret (e.g. *sipring*). However, whether this is so is a question for future research.

The last category, distortions, raised questions for us as we analyzed the data. The use of [r] or [ɾ] at the beginning of a syllable is unlikely to be interpreted as any category but English /ɹ/, but the large numbers of [r] and [ɾ] pronunciations in the data suggests that these types of segmental deviations may be important for comprehensibility. In English, /ɹ/ is a common segment, and when nearly half of /ɹ/ pronunciations are pronounced as a non-English sound, there may be an effect on how much listeners have to work to process speech. In addition, /ɹ/ in English also occurs postvocalically, and when this more vocalic realization of /ɹ/ is pronounced as /r/, listeners may again have to work harder to interpret the consonantal realization of the trill as the vocalic allophone of /ɹ/. The flap, or [ɾ], variant, may pattern differently from the trill. The flap occurs in North American English as an intervocalic allophone of /t/ or /d/, and is not likely to be interpreted as /ɹ/. Because of this, it may be that an intervocalic flap will be misinterpreted (e.g. *carry* heard as *catty*). All of these are only hypotheses, and would require studies to test whether such distortions are merely a marker of accentedness and are thus relatively unimportant, or whether the pronunciation impacts comprehensibility. Because of the frequency of [r] in the corpus, these questions may reveal interesting findings about whether markers of accent can also be important for comprehensibility.

2 Priorities

Another question implicit in this study is about priorities for segmental instruction, and the extent to which the systematic analysis of errors in a corpus is valuable for setting priorities. In other words, which errors should be given priority for advanced Arabic learners of English? This question takes into account issues of frequency and functional load for contrasting errors. For sounds that do not fit measures of functional load, such as deletions and insertions, new measures are proposed. In all cases, the priorities are for persistent errors and are most appropriate for advanced pronunciation learners. It may be that Arabic learners of different proficiency levels would show a different distribution of errors or that a different spoken register would give results different from a reading task.

Table 6 is organized to prioritize frequent errors that are higher in functional load (7–10), followed by errors that change syllable structure because of their ability to mask expected phonetic identity, followed by frequent errors that are salient, followed by frequent errors that are low to moderate in functional load (1–6). One error is placed in the last group despite its FL of 7 (/z–ð/). This was done because of its frequency in unstressed function words and its variability even in L1 varieties of English.

Table 6. Teaching priorities for advanced proficiency Arabic English speakers.

Error	Frequency	Functional load (FL)	Priority
Likely priorities:			
/p/–/b/	27.8% /p/, 13.0% /b/	10 /p/–/b/	Two-way problem
Monophthong for diphthong	15.2% /eɪ/ 10.6% /ɛ/ 27.5% /oʊ/ 7.4% /ɔ/	9 /eɪ/–/ɛ/ 10 /oʊ/–/ɔ/	Only diphthongs to show problems
/ɛ/–/ɪ/	10.6% /ɛ/ 5.7% /ɪ/	9 /ɛ/–/ɪ/	Two-way confusion
/z/	23.0%	8 /z/–/s/	Word endings
/v/	12.0%	7 /v/–/f/	/v/ is the problem
Possible priorities:			
/ŋ/	27.0% [ŋg] 12.8% [ŋk]	No FL for insertions	Almost all final /ŋ/ have insertions
Vowel insertion in consonant clusters	.55% of total phones	No FL for insertions	93.7% of all vowel insertions were found in clusters
Deletions	1.11% of total phones	No FL for deletions	
Unknown priorities:			
Wrong variants of <r>	35.1% 22.0% ER	No FL available	Highly salient error
Unlikely priorities:			
/ð/	21.7%	7 /z–ð/ 5 /d–ð/	Potential problems
/ɑ/	22.2%	6 /ɑ/–/oʊ/	Likely spelling issue
/θ/	16.8%	5 /s–θ/	Potential problem
/dʒ/	40.1%	NA /dʒ/–/ʒ/ 3 /dʒ/–/tʃ/	Common but low FL
/ʊ/	7.8% /ʊ/ 7.0% /u/	3 /ʊ/–/u/	/ʊ/ was highly infrequent and /ʊ/–/u/ was low FL
/ʒ/	15% /ʒ/–/ʃ/	NA	5 errors of 20 total tokens

VI Implications

The fundamental goal of this study was to examine how well a systematic collection of phonetically-annotated spoken language could identify persistent segmental errors. The study examined advanced Arabic speakers of English, comparing error patterns from their oral reading production to the error analysis of various experts. As expected, this data-driven approach showed that the experts missed a number of common errors while accurately identifying others. Experts also sometimes overgeneralized error patterns (such as /v/–/f/, where only /v/ was a common problem). Detailed descriptions of the patterns were also easier to show with even a modest dataset of read speech. Overall, the results of this study indicate that the use of a phonetically-annotated spoken corpus/dataset is promising for identifying targets for pronunciation teaching. A corpus that represents different proficiency levels and spoken language tasks would make the analysis better, of course, but the modest collection of data represented here revealed patterns that we had not found described by expert analyses.

To date, phonetically-annotated corpora are rare because of the amount of work involved in such annotations. While it would be valuable to have a corpus that is both authentic and connected to a specific discourse domain (Biber, 2019), some studies that have done this (e.g. McAndrews & Thomson, 2017) have based their conclusions on smaller amounts of speech (about 10 minutes) than we were able to use (about 40 minutes). Depending on the goals of the research, larger amounts of speech may better represent common challenges faced by groups of learners. In other words, the question of whether L2 Arctic counts or does not count as a corpus is beside the point from a pronunciation teaching point of view. A greater range of L2 sounds can be efficiently elicited from reading aloud than from an equivalent amount of free speech, and the limitations due to the complications of spelling–sound correspondences may even tell us more about how well L2 speakers navigate the writing–speaking divide in L2 learning.

This study has implications for Arabic L1 learners of English in the classroom. One implication is found in the frequency of errors even at an advanced level. Over 10% of all phones in the corpus were clearly inaccurate, a surprisingly high number for such advanced speakers. This suggests that judgments of comprehensibility and accentedness may be affected by the number of errors not only in a single sentence but also in speech in discourse. Munro and Derwing (2006) found that two high FL errors had a cumulative effect on judgments of accentedness, but they stopped at two, approximately half of what our sentences had on average. Examining such judgments with higher numbers of errors may shed light on what happens when errors are compounded.

The results may also be an effect of the reading task, and it would be valuable to elicit comparable spontaneous speech samples to determine the extent to which sound–spelling correspondences affect pronunciation. Given the possibility that errors were connected to sound–spelling correspondences, teachers may find that it is useful to use the International Phonetic Alphabet (IPA) to raise students' phonological awareness of English. Recent investigations in classrooms of English as a second language (ESL) and English as a foreign language (EFL) have often reported that improved phonological awareness in English is beneficial to Arabic L1 learners (e.g. Hago & Khan, 2015; Rajab, 2013). One method used in recent intervention studies for Arabic L1 learners is the teaching of the phonetic alphabet (Mirza, 2015; Rajab, 2013). For example, Rajab (2013) suggested that the teaching of IPA may help raise Saudi students' phonological awareness of English, which in turn improves their English speaking (and presumably, pronunciation) skills.

The use of IPA may be combined with other strategies, such as minimal pair training and English phonotactics instruction. Recent intervention studies have suggested that minimal pair training may be helpful for Arabic L1 learners of English (Mirza, 2015). For example, Altamimi (2015) has found that 2nd intermediate grade Arab speakers were able to improve their pronunciation of English /p/, /ʒ/, /v/, /tʃ/ and /ŋ/ after four weeks of minimal pair training. Another potentially useful strategy is to teach students the phonotactics of English. Al-Jasser (2008) reported that Saudi students improved their English word recognition after eight weeks of explicit teaching of English phonotactics (legal vs. illegal clusters in English). The present finding that many clusters had inserted vowels suggests that students may benefit from explicit instruction of English phonotactics, and they may benefit even more if the instruction uses IPA. These studies, however, focus on accuracy not intelligibility, the highest standard for evaluating pronunciation improvement (Derwing & Thomson, 2015).

A limitation was our lack of attention to suprasegmentals. This was deliberate. Annotations of segmentals are straightforward. Prosodically annotated corpora, although more common, depend upon particular models of prosody, and they were not part of L2 Arctic. In addition, segmentals are worthy of study apart from prosody. They continue to dominate the teaching of pronunciation for many teachers. We know that they are implicated in loss of intelligibility (Im & Levis, 2015; Jenkins, 2000; Zielinski, 2008), and that it is important to know how to determine priorities for segmentals in pronunciation teaching (Munro, Derwing & Thomson, 2015). An accurate identification of segmental errors is also essential in understanding the extent to which such errors are common for groups of learners. This knowledge is also important to identify priorities based on functional load considerations. Such were some of our goals in this study. Suprasegmentals are equally important. However, they were not our target in this study.

The findings of this study can be used for future developments of computer-assisted pronunciation training (CAPT) tools for L1 Arabic learners of English. Information on persistent errors of L1 Arabic learners can inform developers of mispronunciation detection systems to facilitate advancements in artificial intelligence and machine learning for L2 English learners of Arabic L1 background. This can be further used to develop L2 pronunciation learning software that can enhance and complement classroom instruction of English to L1 Arabic learners. These findings can also help improve the mathematical models used to develop automated speech recognition (ASR) toolkits, as such toolkits can use information from our corpus to decide on correctness or incorrectness of the phonemes (Maqsood et al., 2017).

VII Conclusions

Fundamentally, we hoped to demonstrate in this study that the use of a phonetically-annotated corpus with a systematic analysis of highly proficient speakers' production would reveal patterns about pronunciation priorities that expert views had missed or overstated. In this, we were not disappointed. The corpus revealed certain errors that we had not seen from the experts, confirmed others, and raised questions about whether other errors were frequent enough to be taught. We chose to look at Arabic speakers because their likely English pronunciation errors, especially segmental errors, are well-studied and have been described by a variety of experts. However, the errors had not been prioritized in regard to either frequency, functional load, or other criteria. The use of a corpus allowed us to look at the frequency of persistent errors, and modified by the use of functional load, where possible, allowed us to prioritize errors. In addition, the corpus also allowed us to identify errors that were not discussed by the experts. Ultimately, these findings should be confirmed and modified by listener judgments of comprehensibility, intelligibility and accentedness. The errors identified in the corpus offer a direction for collecting such listener judgments.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by NSF awards 1619212 and 1623750.

ORCID iD

Ivana Rehman https://orcid.org/0000-0002-8590-5401

References

Alezetes, E.D. (2007). *A markedness approach to epenthesis in Arabic Speakers' L2 English*. Unpublished masters thesis, University of Montana, Missouri, MT, USA.

Al-Jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, *36*, 94–106.

Al-Mutawa, N., & Kailani, T. (1989). *Methods of teaching English to Arab students*. Harlow: Longman.

Altamimi, A.K. (2015). Improving English pronunciation among Arabic EFL school-age students using minimal pairs. Unpublished masters thesis, SUNY Fredonia, Fredonia, NY, USA.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529–555.

Avery, P., & S. Ehrlich (Eds.). (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.

Best, C.T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O.S., & M. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 13–34). Amsterdam: John Benjamins.

Biber, D. (2019). Corpus analysis of spoken discourse: Research findings, prospects, implications for teaching. Plenary speech given at Pronunciation in Second Language Learning and Teaching conference, September 2019, Northern Arizona University, Flagstaff, AZ, USA.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, *22*, 593–606.

Catford, J.C. (1987). Phonetics and the teaching of pronunciation. In Morley, J. (Ed.), *Current perspectives on pronunciation* (pp. 87–100). Alexandria, VA: TESOL Press.

Catford, J.C. (1988). Functional load and diachronic phonology. In Tobin, Y. (Ed.), *The Prague School and its legacy* (pp. 3–19). Amsterdam: John Benjamins.

Celce-Murcia, M., Brinton, D.M., & Goodwin, J.M. (2010). *Teaching pronunciation: A course book and reference guide*. Cambridge: Cambridge University Press.
Chan, A.Y., & Li, D.C. (2000). English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. *Language, Culture and Curriculum*, *13*, 67–85.
Čéplö, S., Bátora, J., Benkato, A., Milička, J., Pereira, C., & Zemánek, P. (2016) Mutual intelligibility of spoken Maltese, Libyan Arabic, and Tunisian Arabic functionally tested: A pilot study. *Folia Linguistica*, *50*, 583–628.
Derwing, T., & Thomson, R. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*, 326–344.
Deterding, D. (2013). *Misunderstandings in English as a lingua franca: An analysis of ELF interactions in South-East Asia*. Berlin: Walter de Gruyter.
Fox, R.A., Jacewicz, E., Eckman, F., Iverson, G.K., & Lee, S.A. (2009). Perception versus production in Korean L2 acquisition of English sibilant fricatives. In: *Current issues in unity and diversity of languages: Papers selected from CIL 18*, *held at Korea University, Seoul, 21–26 July 2008*. Seoul: The Linguistic Society of Korea, pp. 2662–2680.
Gilner, L., & Morales, F. (2010). Functional load: Transcription and analysis of the 10,000 most frequent words in spoken English. *The Buckingham Journal of Language and Linguistics*, *3*, 135–162.
Hago, O., & Khan, W. (2015). The pronunciation problems faced by Saudi EFL learners at secondary schools. *Education and Linguistics Research*, *1*, 85–99.
Hockett, C.F. (1966). *The quantification of functional load: A linguistic problem*. ERIC. Available at: https://files.eric.ed.gov/fulltext/ED011649.pdf (accessed May 2020).
Im, J., & Levis, J. (2015). Judgments of non-standard segmental sounds and international teaching assistants' spoken proficiency levels. In Gorsuch, G. (Ed.), *Talking matters: Research on talk and communication of international teaching assistants* (pp. 113–142). Stillwater, OK: New Forums Press.
Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
Kharma, N., & Hajjaj, A. (1989). *Errors in English among Arabic speakers: Analysis and remedy*. Harlow: Longman.
Levis, J.M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 369–377.
Levis, J.M. (2018). *Intelligibility, oral communication and the teaching of pronunciation*. Cambridge: Cambridge University Press.
Maqsood, M., Habib, H.A., Anwar, S.M., Ghazanfar, M.A., & Nawaz, T. (2017). A comparative study of classifier based mispronunciation detection system for confusing Arabic phoneme pairs. *Nucleus*, *54*, 114–120.
McAndrews, M.M., & Thomson, R.I. (2017). Establishing an empirical basis for priorities in pronunciation teaching. *Journal of Second Language Pronunciation*, *3*, 267–287.
Mirza, H.S. (2015). ESL and EFL learners improve differently in pronunciation: The case of Lebanon. *Procedia-Social and Behavioral Sciences*, *199*, 486–495.
Mitchell, T.F., & El-Hassan, S. (1989). *English pronunciation for Arabic speakers*. Harlow: Longman.
Munro, M.J. (2018). How well can we predict second language learners' pronunciation difficulties? *CATESOL Journal*, *30*, 267–281.
Munro, M.J., & Derwing, T.M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*, 520–531.
Munro, M.J., Derwing, T.M., & Thomson, R.I. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teaching*, *53*, 39–60.

Nilsen, D.L., & Nilsen, A.P. (1971). *Pronunciation contrasts in English*. New York: Regents.

Nguyen, N. (2008). Interlanguage phonology and the pronunciation of English final consonant clusters by native speakers of Vietnamese. Unpublished masters thesis, Ohio University, Athens, OH, USA.

No, G. (1997). Developmental and transfer factors in English pronunciation of Korean elementary school and college students. *English Teaching*, *52*, 21–39.

O'Brien, M.G., Derwing, T.M., Cucchiarini, C. et al. (2018). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, *4*, 182–207.

Prator, C., & Robinett, B. (1985). *Manual of American English pronunciation*. 4th edition. New York: Holt, Rinehart & Winston.

Rajab, H. (2013). Developing speaking and writing skills of L1 Arabic EFL learners through teaching of IPA phonetic codes. *Theory and Practice in Language Studies*, *3*, 653–660.

Smith, B. (2001). Arabic speakers. In Swan, M., & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (pp. 195–213). Cambridge: Cambridge University Press.

Surendran, D., & Niyogi, P. (2003). Measuring the usefulness (functional load) of phonological contrasts. Technical Report TR-2003-12. Chicago, IL: Department of Computer Science, University of Chicago.

Surendran, D., & Niyogi, P. (2006). Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In Thomsen, O.N. (Ed.), *Competing models of linguistic change: Evolution and beyond* (pp. 43–58). Amsterdam: John Benjamins.

Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems*. 2nd edition. Cambridge: Cambridge University Press.

Uzun, T. (2019). Pronunciation and intelligibility of Turkish native speakers learning English. Unpublished PhD thesis, Ankara University, Ankara, Turkey.

Van Doremalen, J., Cucchiarini, C., & Strik, H. (2009). Optimizing automatic speech recognition for low-proficient non-native speakers. *EURASIP Journal on Audio, Speech, and Music Processing*, *2010*, 973954.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, *4*, 123–130.

Zhao, G., Sonsaat, S., Silpachai, A. et al. (2018). L2 ARCTIC: A non-native English speech corpus. *English Publications*, *226*. Available at https://lib.dr.iastate.edu/engl_pubs/226 (accessed May 2020).

Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. Unpublished PhD thesis, MacQuarie University, Sydney, Australia.

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, *36*, 69–84.

Appendix 1. List of ARPAbet symbols and corresponding IPA symbols and examples.

Number	ARPAbet	IPA	Example
Consonants:			
1	B	b	**b**ow
2	CH	tʃ	**ch**eck
3	D	d	**d**eed
4	DH	ð	**th**is
5	F	f	**f**all
6	G	g	**g**ood
7	HH	h	**h**eed
8	JH	dʒ	**j**eep
9	K	k	**k**ing
10	L	l	**l**ike
11	M	m	**m**ine
12	N	n	**n**ot
13	NG	ŋ	wi**ng**
14	P	p	**p**eep
15	R	ɹ	**r**ight
16	S	s	**s**ore
17	SH	ʃ	**sh**oe
18	T	t	**t**en
19	TH	θ	**th**umb
20	V	v	**v**ain
21	W	w, ʍ	**w**oe
22	Y	j	**y**es
23	Z	z	**z**ebra
24	ZH	ʒ	vi**s**ion
Vowels:			
1	AA	ɑ	n**o**t, t**al**k
2	AE	æ	h**a**t
3	AH	ʌ	c**u**t
4	AH0	ə	comm**a**
5	AO	ɔ	c**o**re
6	AW	aʊ	sh**ou**t
7	AY	aɪ	m**i**ght
8	EH	ɛ	w**e**t
9	ER	əɹ	f**ir**st
10	EY	eɪ	w**ai**t
11	IH	ɪ	p**i**t
12	IY	i	m**ee**t
13	OW	oʊ	g**oa**t
14	OY	ɔɪ	j**oy**
15	UH	ʊ	h**oo**k
16	UW	u	sh**oo**t
Other:			
	*		R* = (trilled) /r/

Note. * indicates a nonnative variant of a native category