Review of dissertation

The Perception and Production of American English Sounds by Palestinian Arabic Adolescents

Bashar M. M. Farran

Supervisors: Prof. Dr. Vincent J. van Heuven, Dr. Ildikó Hortobágyi

Reviewer: Szilárd Szentgyörgyi

Bashar Farran's work is a well-designed, logically structured PhD dissertation supported by a large number of experimental data and a thoroughly researched overview of the relevant literature. I have to say up front that I consider the dissertation valuable and suggest that the candidate should be awarded the PhD degree after a successful defence of his dissertation.

The dissertation examined the perception and production of American English (AE) monophthongs by Palestinian Arabic EFL learners. The structure of the paper was as follows: Chapter 1 described Palestinian Arabic (PA) learners of English as a Foreign Language to identify their EFL status together with the pedagogical state of affairs. Also, the main research questions and the broader hypotheses were presented.

Chapter 2 presented the literature review made up of two subsections. The first of these consisted of the phonetic overview of the languages dealt with in the dissertation by comparing PA features with the phonotactic characteristics of AE. The second subsection in Chapter 2 provided an overview of the most updated and well-known L2 learning models, approaches, and their most recent versions. On the basis of this, the central methodology of the dissertation was assimilation, identification, and production testing in three consecutive studies done with non-native PA EFL learners and native AE speakers. These studies are presented in Chapters 3-5 one by one.

The author notes on page 12 that "*The double marking of the contrast (in duration and in vowel quality) would be an argument to consider the contrast as a tense-lax opposition rather than just a length opposition.*" There are several notes to be made here. First of all, the fact that long-short vowel pairs do not only differ in length but also in quality is quite common among languages. English, regardless of whether you look at Received Pronunciation or General American varieties, for instance, does not have one single short-long vowel pair that has exactly the same quality, which is clearly demonstrated by the fact that different IPA symbols are used to denote members of such long-short vowel pairs, e.g. /iː-ɪ/, /uː-ʊ/, /ɑː-ɒ/, etc. In other languages like Hungarian, the situation is more similar as the quality difference is not large

enough to encode it into using different IPA symbols but there is a tendency to produce long vowels pushed more towards the peripheries of the vowel space than short vowels. Thus, similarly to Arabic, Hungarian /iː/ is produced with a somewhat higher and more fronted tongues position than short /i/ while /uː/ is produced with a higher and somewhat more back tongue position than short /u/. Thus, what we see in Arabic in this aspect does not seem to be like the English case but very much like the Hungarian case, so it could be concluded that the more extreme nature of long vowels is an automatic phonetic characteristic and not something to be encoded into their phonological representation. A look at the literature in the long-short vowels of other languages should have revealed this. Thus, referring to an underlying tense vs. lax difference seems an overshot in explaining this difference. As a result, the tense-lax difference should only be used in a strictly phonetic sense and not when referring to underlyin, i.e. phonemic, differences between vowel segments.

The other note that needs to be made here is that when comparing English with Arabic, one has to take into consideration the fact that vowel length in the former language is extremely unstable: all underlyingly long vowel phonemes – long monophthongs, as well as long diphthongs and triphthongs – tend to shorten significantly whenever followed by fortis – i.e. phonologically voiceless – consonants as in *seat* /siːt/-[sit], *light* /laɪt/-[laɪˑt], *house* /haʊs/-[haʊˑs], *boot* /buːt/-[but], *fake* /feɪk/-[feɪˑk]. Thus, if one examines how well Arabic speakers do when trying to imitate native English vowel quality and quantity, one cannot put this issue aside as in the dataset there surely are words that will pose a problem for Arabic speakers because of the reasons described above.

As for the consonant system, there are some inconsistencies as well: the author lists the members of the consonant system including the glottal stop [ʔ] but does not do so with the voiced alveolar flap/tap [ɾ], the voiceless variant of the labio-velar glide [ʍ] or the dark variant of the alveolar lateral [ɬ]. One may argue that the flap does not appear in monosyllabic words, yet all the others do. This brings up a more general issue that has already been discussed with the author concerning earlier presentations of the research, namely whether the discussion is focussing on the actual physical reality of the speech sounds, i.e. phonetics, or their abstract mental representation, i.e. phonology. At most places in the dissertation it seems that it is the physical reality of speech sounds, that is phonetic entities, that the research focusses on – this is exactly what the author explicitly states at many places concerning the tense-lax distinction, for instance. However, at other places, it seems that we are slipping into a discussion of the phonological system of the two languages. This is exactly the case when the author discusses the sound systems of the two languages: the tables only contain what are traditionally

considered members of the phoneme systems of the two languages. Some allophones – like the glottal stop in the case of the consonants of English or the diphthongization of /e, o/ - are mentioned but others are completely left out of the discussion. I think that it should have been made entirely clear whether it is the phonological system and the behaviour/distribution of some sound segments that the investigation is focussing on, i.e. it is a matter of phonology, or the physical – articulatory, acoustic, or auditory – reality of the sound segments that the dissertation is concerned with, i.e. it is a matter of phonetics. If both, they should have been treated clearly separated from each other.

As for some minor problems, the following may be listed in chapter 2:

- on page 21 above Figure 2.2. there is a terminological error: "First, vowels are described according to backness (front, middle, or back)", where *middle* should be replaced by *central* when referring to the horizontal dimension of the position of the tongue;
- at the bottom of Table 2.5, the /aː/ symbol for the low back unrounded vowel should be replaced by /ɑː/ as the former only occurs in some regional varieties like Boston English but it is not a monophthong in the variety described.
- page 22: AE /ɹ/ is often argued to be retroflex, i.e. it is the IPA symbol /ɻ/ and not /ɹ/ that should be used;
- page 24: the categorization of obstruents into voiced and voiceless subclasses and claiming that the sonorants /m, n, ŋ, l, r, w, j/ are all voiced has the following problems:
  - even though /h/ has already been classified as a fricative, it is not classified into either category, obstruent or sonorant, as it does not fit either from the point of view of voicing – the author should either claim that it is an obstruent fricative without a voiced counterpart or claim that it is a voiceless sonorant glide (probably the better is the more widely accepted claim these days);
  - the voicing of obstruents is far from stable as the phonologically voiced (lenis) obstruents may all (partially or fully) lose their voiced quality and may surface in pronunciation as (at least partly) voiceless as a result of devoicing in initial, final positions or next to a phonologically voiceless (fortis) segment, e.g. Bob [b̥ɑb̥], obtain [əb̥ˈteɪn]
- page 25: "in CV syllables, any consonant can occupy the onset of the syllable except for /ʒ/ and /ŋ/" – the former consonant does occur in a limited set of words in a simple onset, e.g. *genre*;

- page 25: “Affricates, voiced fricatives, and semi-vowels /h, j/ are never part of any onset cluster” – this is not entirely true as in words like *hue* /hjuː/, *humid* /ˈhjuːmɪd/ you have both semi-vowels appearing in combination.
- page 25: “Here, the choice of C1 is fairly unrestricted: it can be a plosive (pure, beauty, dune, tune, cute but not /gj/), nasal (mew, new), fricative (few, view, suit but not /zj/), liquid /l/ (lute)” – note that in the latter types of word as in *lute*, the /j/ will always be dropped in American English as the obligatory Yod-dropping rule applies after all coronal consonants including /l/;
- page 25: /streŋgθ/ should have a /g/ or /k/ segment after the nasal;
- page 25: “The longer sequences are always a proper subset of a shorter sequence plus one extra C” – the principle is stricter than this: in a $C_1C_2C_3$ onset sequence, both C_1C_2 and C_2C_3 should make a well-formed two-member onset.
- page 26: “At least one syllable in a word receives secondary stress if it precedes the primary stress in Polysyllabic words;” – this is not entirely correct since if there is only one syllable before the primary stressed syllable in the word, e.g. *ahead* [əˈhed], then it will normally not receive secondary stress as it would create a stress clash, i.e. two subsequent syllables carrying major stresses. I’d change the wording to “if a word has at least two syllables before the primary stressed one, then one of these must be secondary stressed”;
- page 26: “Perceiving stress is said to be influenced by a number of factors, e.g., increased segment length, loudness, pitch variation, and quality (Fry, 1958; 1965; Van Heuven, 2018), although a syllable can be made prominent even with only one or two factors present.” English so-called major stressed syllables – i.e. the ones that carry primary or secondary stress – tend to be rhythmically prominent as they are louder than non-major stressed syllables. Primary and secondary stressed syllables are distinguished by the former having higher pitch than the latter. Non-major stressed syllables – tertiary stressed and unstressed syllables – are distinguished from each other by the former not having reduced but full vowels (no centralization, no shortening) while the latter have reduced vowels only (centralized and shortened);
- page 27, Table 2.9: Why do you expect /ɪ/ to be confused with /ɛ/ but not with /e/? Articulatorily the latter is much closer to /ɪ/ than the former.
- page 28: „the realization of /i:/ and /u:/ will be reasonable since these vowels will have a sufficiently peripheral quality and will be pronounced long” – it is mostly true but because of the process of Pre-Fortis Clipping, these phonemically long vowels will often

surface just as short as a real short vowel if followed by a fortis (phonemically voiceless) consonant, e.g. /siːt/-[sit], /buːt/-[but]. You may surely expect some confusion in their perception in these cases. Consequently, if you test PA speakers' perception purely based on examples where the length of the English vowels is always stable, i.e. they are always fully long, then it is a distortion of the influences that the same speakers will receive when faced with real life English data where some of the time the same (phonologically long) vowels will be shortened, i.e. the length cue will not be available to base their judgements on. Do you think that a similar experiment with balanced environments for the vowels – i.e. half the tokens before voiced, the other half before voiceless segments – would give you a more realistic idea of what PA speakers' identification skills for these vowels are really like?

- page 29. Table 2.10: What do you expect to happen to some English consonant allophones: e.g. the devoiced versions of /ɹ, l, j, w/ after aspirated voiceless stops – e.g. /twin/-[tw̥ɪn], /preɪ/-[pr̥eɪ] – or the velarized version of /l/ - called a Dark-L – in syllable rhymes as in /fiːld/-[fiːɬd] or /pʌls/-[pʌɬs]? Since PA does not have any of these consonants, you may as well expect quite an amount of confusion concerning their identification as well as production.

Chapter 3 discusses PA EFL learners' assimilation patterns of AE L2 vowels to PA native vowels by applying a method developed within the Perceptual Assimilation Model (PAM). The results presented show great confusion of the AE monophthongs by PA EFL learners, which in turn suggest that PA learners of AE will be likely to make perception errors – and possibly production errors, too – as laid out in Chapters 4 and 5. Since one of the main findings is that PA leaners of AE heavily rely on vowel length to identify vowels, it would have been interesting to see how the subjects react to carrier words in which the phonologically long vowels shorten as that would have been a nice control case to see how they interpret pure quality difference between long /iː/ in *heed* [hiːd] vas a shortened /iː/ in *heat* [hiˑt]. It was to be expected that PA speakers will rely on vowel length as there is not much of quality difference in the PA vowel system and that is exactly why looking at the lack of such length difference would have been interesting to see.

Minor questions and notes for Chapter 3:

- page 65: there may be some possibility for misinterpreting your informants' reactions in the experiment here. "*The Arabic responses read: /saːr/ 'walked', /sar/ 'to rejoice',*

/su:r/ 'outside-/outer wall', /sur/ 'rejoiced', /si:r/ 'walk imp.', and /sir/ 'secret'". Since you mentioned that speakers of different varieties of PA will have different allophones for the vowel phonemes, how did you exclude the possibility of a test subject that pronounces a word like /sir/ with an allophone for the vowel like [e] would correctly identify the vowel in the sample word "hayed" as [e] which corresponds to their pronunciation of /sir/. In this case, you may identify it as a misperceived vowel even though the test subject's perception was the closest possible to the input.

- page 70: "*AE /a/ is in between PA /a:/ and /u:/, but closer to the former than to the latter*", In what way is /ɑ/ between /aː/ and /uː/? Articulatorily, it is surely not between them at least not at an equal distance from both of them, it is much closer to /ɑː/ than to /uː/.

Chapter 4 introduces a perception experiment with vowels synthesized in Praat, which had to be identified as tokens of the AE monophthongs by native AE and PA EFL listeners alike to help display differences in the mental representation of the vowel space between the two types of speaker. The results of Chapter 4 show that the agreement of PA perception choices with the those of native AE speakers was not very strong as it did not reach 34%. The only exceptions were the high long vowels /iː, uː/ as they are almost perfect matches between the two languages. PA EFL learners perceived AE vowel duration in a way that is highly congruent with the native norms but the correlation for perceiving the quality of the vowels is rather weak between natives PA and native AE speakers. These results nicely correspond to the PAM predictions from the previous chapter and support the claim that duration is the primary source of differentiation between AE vowels. Again, had the author also tested the perception of long AE vowels before voiceless consonants, we could have received some more insight into what degree PA speakers rely on the length of AE vowels in perception.

Minor questions and notes for Chapter 4:

- page 84: "The synthesized reference set of tokens contains a labial onset and coda, /m/ and /f/, respectively, and the synthesized nucleus in between them yielding a /m_f/ carrier." Did you consider the shortening effect of the voiceless fricative on the preceding long vowels, were the (phonologically) long vowels shortened accordingly in the stimuli?
- page 86: "The response buttons contain well-known and meaningful monosyllabic (real) English words written in English orthography in a fixed /CVl/ carrier covering the 11 pure vowels of AE." The problem is that in this context – i.e. before a dark,

velarized [ɬ] – vowels tend to reduce the height of the tongue and as a result, their F1 values will be different from those of the same vowels before other consonants, which may influence participants' judgements.

- page 98: "Since the native listeners deviate from the modal response per vowel category in 40% of the cases, there is substantial confusion (or disagreement) even for the native listeners in Figure 4.8A." This may well be the result of the differing native dialects spoken by the AE speakers.
- page 102: "Figure 4.7 shows that duration plays a much more prominent role in the perceptual representation of the AE vowels for the EFL learners than for the native listeners." Since you only used stimulus data not showing any long vowel shortening examples, it follows that the real situation is much worse when PA speakers try to identify AE vowels as in a great percentage of the cases – i.e. when a voiceless consonant follows the (phonologically) long vowel – the length of the vowel as an indicator will not be available for EFL learners, i.e. you may expect much worse percentages on a test including vowel shortening cases.
- page 102: "These prominent confusions, however, never involved the tense-lax contrast". Of course, since confusing examples – e.g. *court, take, boot, seat,* etc. – were not used in the stimulus set.
- page 113 H1: "Less problematic are the high vowels with tense-lax contrasts between /i-ɪ/ and /u-ʊ/, which will still be differentiated based on duration differences rather than a difference in vowel quality." But only if you exclude long vowels followed by voiceless consonants.

In Chapter 5, the author presents the results of yet another experiment, this time concerned with the production of AE vowels by PA EFL learners to be compared with the production data of the same vowels by native AE speakers. According to the findings concerning PA EFL learners' speech production, the formant values of vowels adjacent to each other in the vowel space often overlap with each other, and AE vowels without a corresponding PA vowel show overlapping production.

Chapter 6 is a synthesis of the three experimental research studies and compares the PA EFL learners' perception and production in an integrated fashion to see if a good perceptual representation is necessary for the good pronunciation of an AE target sound – interestingly, no

strong correlation was found between learners' perception and production of AE vowels at the individual level.

Finally, chapter 7 is concerned with the pedagogical implications as well as the limitations of the study and possible directions for future research in the topic.

Finally, I conclude and repeat that Bashar Farran's PhD dissertation is valuable scientific work, which meets all the requirements concerning form and content for doctoral dissertations in the Multilingualism Doctoral School in Linguistics at the University of Pannonia and I suggest that the author should be awarded the PhD degree if his defence procedure is successful in all respects.

August 16th, 2022. Veszprém

Szilárd Szentgyörgyi
University of Pannonia