Perceptual representation and production of American English simplex vowels by adolescent Palestinian-Arabic learners

Vincent J. van Heuven[1, 2, 3] & Bashar M. M. Farran[1]

[1] Multilingual Doctoral School, University of Pannonia, Veszprém, Hungary
[2] Leiden University Centre for Linguistics, Leiden, The Netherlands
[3] Frisian Academy, Leeuwarden, The Netherlands
v.j.j.p.van.heuven@hum.leidenuniv.nl; bfarran89@gmail.com

ABSTRACT

To establish their perceptual representation of the American English (AE) vowels, 40 adolescent Palestinian-Arabic (3 long + 3 short vowels) learners of English as a foreign language and 20 American native controls identified each of 86 artificially generated vowel sounds (in /mVf/ nonsense items) as one of the 11 simplex vowels of AE (forced choice). F1 was varied in 7 steps of 1 Bark, F2 in 9 steps (20 impossible F1-F2 combinations were excluded). Vowel durations were either 200 or 300 ms. Results reveal large deviation of vowel centroids in the L2 responses (from L1 control centroids), and much greater overlap of spreading ellipses. The effect of duration was stronger in L2 than in L1 responses. The perceptual representations matched the locations of the AE vowels in the acoustic vowel diagram based on the same learners' production data, suggesting a close link between vowel perception and production.

Keywords: Perceptual representation, synthesized vowels, L1 interference, Arabic, American English.

1. INTRODUCTION

1.1. Importance of perceptual studies for L2 learning

When comparing the vowel systems of native (L1) and foreign (L2) languages, the usual procedure is to record speakers of the language variety of interest and then measure the lowest two to four resonances in the speaker's vocal tract as an indication of how each vowel is pronounced. The lowest resonance (formant F1), corresponds to the openness of the vowel, while the second-lowest formant (F2) is an indication of backness and lip rounding [1, 2]. Plotting the F1 and F2 values as vowel coordinates in a two-dimensional map then gives a good impression of the general organization of the vowel system. The mean (also called centroid) of the dispersion cloud of each vowel is taken as the most representative ('prototypical') realization of the particular vowel type. Vowel duration is often added as a third parameter to define the vowel space of the language (variety).

By comparing the target system with the learner's L1, differences and similarities in the organization of the respective vowel systems can be illustrated, potential learning problems can be identified, and instructions can be formulated inform the learner how to modify their native vowel category so as to articulate a more authentic vowel in L2 (e.g., [3]).

It is insufficiently realized, however, that studying the production of the vowel systems per se does not reveal the full organization of a vowel system, and – more importantly – does not reveal the (often incorrect) perceptual representation of the vowel system of the L2 in de mind of the learner. What is needed to appreciate the representation of the vowel system in the mind of the learner (and of the native speaker) is a perceptual mapping. Using perceptual techniques allows the researcher to establish so-called trading relationships between the parameters that define the individual listener's vowel space.

The present paper therefore aims to establish the perceptual representation of the vowels of American English (AE) in the minds of adolescent (secondary school) intermediate-level Palestinian-Arabic learners of English as a foreign language (EFL). We compare the perceptual representation with vowel production data from the same speakers. We test the hypothesis that the perceptual representation matches L2 production, and that L2 perception and production deviate in the same way from L1 norm data.

1.2. Vowels of Palestinian Arabic vs American English

The vowel system of Modern Standard Arabic (MSA) is generally analysed with 6 phonemes [4], i.e., corner vowels /i, a, u/ and contrastive length (Figure 1, left). Additionally, Palestinian Arabic (PA) has mid-vowel allophones [e, eː, o, oː], which may affect EFL learning [5: 171, 6: 529, 7: 1, 8]. American English is analyzed with 11 simplex vowels, i.e., 4 lax/short vowels /ɪ, ɛ, ʊ, ʌ/ and 7 tense/long vowels /i, e, æ, ɑ, ɔ, o, u/, of which /e/ and /o/ diphthongize (but not enough to consider them complex vowels [9: 114-115]), with positions in the IPA vowel diagram as in Figure 1, right. The short/lax vowels (joined by the inner quadrilateral in Figure 1) are articulated closer to the center of the vowel space.

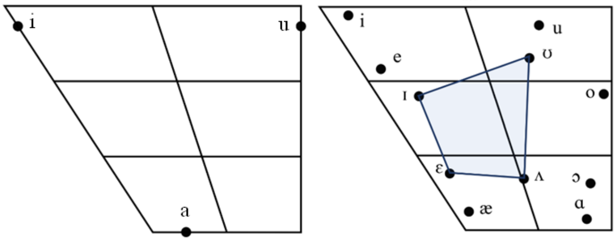

Figure 1: Vowel diagrams for Arabic (left, copied from [4]) and American English (right, after [10]).

2. METHODS

2.1. Stimuli

We used the universal reference set of synthesized vowel tokens proposed by [11] to map out the (oral) monophthongs of any language. The vowel space is defined by three parameters, i.e., vowel height (F1), vowel backness/rounding (F2) and length (vowel duration). The vowels were synthesized in an acoustic vowel triangle defined by the F1-by-F2 coordinates of extreme, prolonged vowel tokens pronounced as /mif/, /maf/ and /muf/, technically not cardinal vowels but contexted tokens which are a close approximation to the cardinals. To ensure auditorily uniform inter-vowel spacing, F1 and F2 stepsizes were always 1 Bark. In the grid, F1 values ran from 2.5 to 8.5 Bark, in 7 steps. F2 values were varied in 9 steps between 6 and 14 Bark. This yields 9 × 7 = 63 vowel spectra in a rectangular matrix (Figure 1). Twenty F1-by-F2 combinations sound inhuman (grey in Figure 2). When these are eliminated, a subset of 43 remain, which were synthesized with (ecologically valid) 200 and 300-ms vowel durations (for background and details, see [11, 12]).

F1			F2 (step number, Hz, Bark)								
			1.	2.	3.	4.	5.	6.	7.	8.	9.
			2357	2031	1746	1497	1278	1086	915	764	628
#	Hz	Bk	14.0	13.0	12.0	11.0	10.0	9.0	8.0	7.0	6.0
1.	237	2.5									
2.	339	3.5									
3.	447	4.5									
4.	565	5.5									
5.	694	6.5									
6.	838	7.5									
7.	998	8.5									

Figure 2: Steady-state F1 and F2 values for 43 reference vowel qualities.

2.2. Participants and procedure

Twenty male and 20 female native speakers of PA were selected as intermediate learners of English. They were secondary school pupils with a mean exposure to (American) English of roughly 6 years in a school setting. These participants listened to the 86 synthesized /mVf/ vowel stimuli over good-quality headphones in individual sessions. The participant saw 11 response buttons on a computer screen, each of which exemplified one of the 11 response vowels by a single keyword, which were supposed to be well-known to the students.

The participants decided, for each /mVf/ token, which of the 11 response vowels came closest to the sound they had just heard, by clicking on the corresponding response button. Stimulus presentation and data collection were controlled by a computer script written for the ExperimentMFC module in the Praat software [13, 14].

In an immediately preceding perceptual assimilation task, the PA participants had been exposed to two tokens of the 11 AE vowels in /hVd/ context spoken by two male native speakers of American English (4 tokens per vowel; see [12] for details).

The procedure was repeated with 20 native listeners of American English, bachelor students of linguistics at the University of Southern California in Los Angeles (7 males) with ages between 19 and 22.[1]

2.3. Collection of production data

The same 40 PA EFL learners produced the 11 AE vowels in everyday keywords as well as in /hVd/ contexts (twice) in a fixed carrier *Now say ... again*. Recordings were made directly (44.1 kHz, 16 bit) on a silent notebook computer in a quiet room, using a Sennheiser PC131 headset microphone. In the results section, we only use the recordings of the /hVd/ words: *heed*, *hid*, *hayed*, *head*, *had*, *hud*, *hod*, *hawed*, *hoad*, *hood*, *who'd*, for details see [14].

3. RESULTS

3.1. Centroids and ellipses in F1-by-F2 vowel space

The results of the identification of the 86 reference vowels in terms of the simplex vowels of AE are shown in Figure 3A-D for the L1 control listeners (top), and for the PA EFL learners (bottom), separately for the 43 short (left, panels A, C) and 43 long (right, B, D) stimulus vowels. The phonetic symbols mark the approximate location of the vowel centroid, while the dispersion ellipses are drawn at ± 1 SD away from the centroid along the first two principal components of the scatter cloud, theoretically including the 46% most typical responses. This presentation of the results facilitates direct comparison of the perceptual representation and the acoustic vowel production. A more traditional presentation of the results in terms of the modal response vowel for each of the stimuli (as used in, e.g., [15, 16]) can be found in [14].

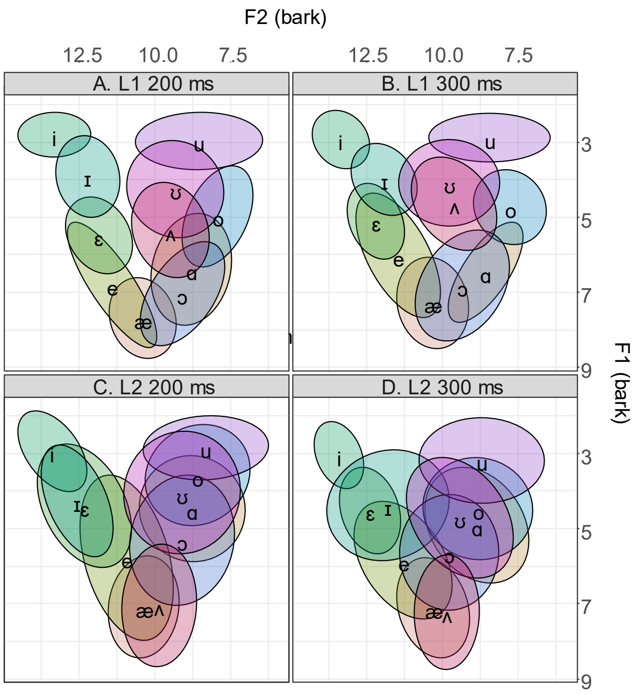

Figure 3. Centroids and dispersion ellipses in F1 by F2 (Bark) plot of American English simplex vowels as identified by American L1 listeners (top panels) and Palestinian Arabic EFL learners (bottom) for 43 synthesized reference vowels with durations of 200 ms (left panels) and 300 ms (right panels). Figure produced with Visible Vowels [17].

The L1 listeners maintain a clear difference between the tense-lax counterparts, /i/-/ɪ/ and /u/-/ʊ/. There is a large distance between the centroids associated with the members of these pairs, and there is only relatively little overlap between the associated spreading ellipses. The centroids of the lax vowels, especially of /ɪ, ʊ, ʌ/, are rather centralized when computed for the stimulus vowels with long duration, but assume more peripheral locations (closer to the tense counterparts) when heard with short duration. This would be a first indication that vowel quality and duration are in a trading relationship in the native speakers' mental representation of the tense-lax vowels. For a phonetically long vowel to be perceived as a lax member of a contrast, it has to be very clearly centralized. When a less centralized vowel is short (enough), it will still be perceived as lax. It is also apparent that the L1 listeners do not maintain distinct vowel categories for pair of tense vowels /ɑ-ɔ/ and for the lax vowels /ʊ-ʌ/. Many speakers, especially in California, no longer distinguish between /ɔ/ and /ɑ/, characteristic of the *cot–caught* merger; also, /ʊ/ is moving towards [ʌ] so that, for example, *book* and *could* start to sound like *buck* and *cud* (e.g., [1]: 212–213).

The L2 participants correctly placed the point vowels /i, u, æ/ at the left-top, right-top and mid-bottom parts of the F1-by-F2-plane, respectively. These vowel qualities were correctly interpreted, presumably because they have suitable near-equivalents in Arabic. Notice that, in the L2 identifications, the centroid for tense /u/ is further back (lower F2 value) than for all other back vowels. In the L1 data, the /u/ centroid is rather more fronted, which corresponds well with the literature on AE.

While the L1 vowel centroids are more or less evenly distributed over the F1-by-F2 space (with the exceptions noted above), substantial clustering of vowels is observed in the L2 identifications. The lax vowels /ɪ, ɛ/ are virtually indistinguishable, as are the members of the /æ, ʌ/ pair. Moreover, the remaining back vowels, /ɑ, ɔ, o, ʊ/ tend to cluster. Typically, the locations of the /ɑ, ɔ/ centroids are too high, while the lax vowel centroids for /ʊ, ʌ/ are located close to the back edge of the vowel space.

3.2. Vowel duration

To obtain an indication of the difference in weight attached to the duration of the stimulus vowel in the perceptual identification, we computed the relative frequencies of the 11 response vowel categories for short (200-ms) and long (300-ms) stimulus vowels (across all 43 synthesized qualities), expressed as a percentage of the total number of responses, i.e., 860 (43 × 20) for L1 responses and 1720 (43 × 40) for L2 responses. These percentages are shown in Figure 4.

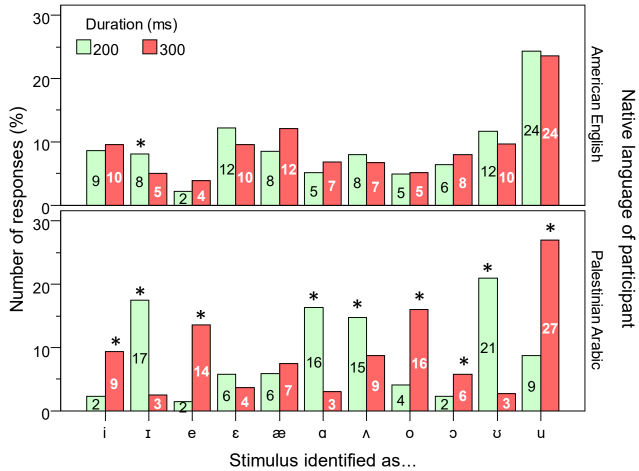

Figure 4. Frequency (%) of 11 response categories for short (200 ms) and long (300 ms) stimulus vowel durations, irrespective of formant structure, for L1 AE (upper panel) and L2 PA listeners (bottom panel). See text for more information.

A significant difference due to stimulus length is seen only for one vowel in the L1 responses, i.e., where the number of /ɪ/ responses is significantly smaller for long stimuli than for short stimuli (all discrepant pairs are marked by an asterisk in Figure 4; binomial test, assuming a 50-50% distribution, $\alpha \leq .05$). For all other vowels, stimulus duration does not affect the response frequencies significantly. For the PA L2 listeners, however, the difference in duration has a significant effect for 8 out of 10 response categories, i.e., for all vowels except /ɛ/ and /æ/.

3.3. Comparison with production data

Figure 5A-D shows the location of the centroids in the F1-by-F2 space found in the perceptual representation in section 3.1, and in the production data recorded from the same PA participants (see section 2.3). The L1 production data have been copied from the literature [18, 19] and, though fully comparable, were produced by different individuals than the participants in the perception experiment.[2]

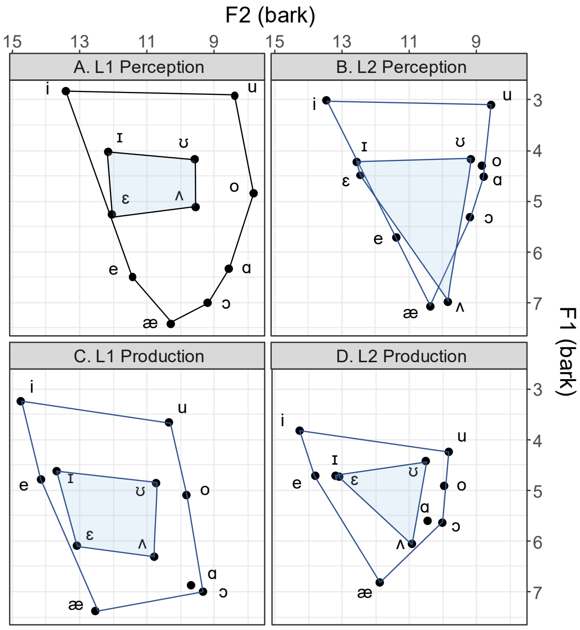

Figure 5. Centroids of 11 AE vowels in F1-by-F2 plane (Bark) found for perceptual representation (top panels) and in speech production (bottom panels) for L1 speakers (left) and for PA EFL learners (right). See text for more information.

The vowel configuration in the production data is shifted away from the origin of the plot relative to the perceptual representation. The latter is based on listeners' responses to a synthesized male voice, whereas the production centroids have been averaged over equal numbers of male and female speakers, which increases the F1 and F2 frequencies due to the shorter female vocal tract. In spite of this, the two L1 configurations (panels A, C) are quite similar. The four short/lax vowels are the corner points of an inner quadrilateral with clear distances from each other and from the nearest long/tense vowels. In the PA EFL configurations, the configurations are rather more triangular. The /ɪ-ɛ/ contrast is absent in both the perceptual representation and in production. Also, The L2 vowel space is compressed in the height dimension in comparison with the L1 configuration, in panel B, and especially in the L2 vowel production (panel D). The compression is most likely caused by the absence of contrastive mid vowels in the EFL learners' native language. Lack of separation between adjacent back vowels is seen in both L2 perception and production.

4. CONCLUSIONS AND DISCUSSION

This study has shown that the perceptual representation of the simplex vowels of American English can be adequately mapped out by having listeners identify a small set of artificial (contexted) vowel sounds varying systematically in F1, F2 and duration. The PA non-native perceptual representation deviates strongly from the L1 configuration, and the non-native categories are less clearly delineated (larger spreading ellipses). The L1 listeners identify the AE vowels mainly on the basis of quality (F1, F2) and tend to ignore vowel duration, as has been shown earlier by artificially lengthening and shortening natural AE vowels [20]. The PA non-native listeners were strongly led by vowel duration, which is contrastive in Arabic but not in English.

As predicted, we found a strong match between the perceptual representation of the AE vowels and the acoustic properties in their production, both for the L1 and for L2 participants at the group level.[3]

The present study was not set up to shed light on the question whether perception leads production or *vice versa*. This would have required following the participants longitudinally. Still, our study shows that production studies should be complemented by research on the perceptual representation of target sounds. The greater reliance on vowel duration by the non-natives cannot be seen in the L2 production data.

The L2 vowel durations (not presented in this paper but see [12, 14]) show less contrast between the AE lax/short vs tense/long vowels than is seen in the L1 production data. Moreover, the L2 speakers produced much shorter vowel durations overall than the L1 speakers, most likely because the L2 speakers failed to apply vowel lengthening before the voiced coda consonant in the /hVd/ words. Similar shorter overall vowel duration in L2 AE speech has been observed for EFL learners whose L1 is Javanese or Sundanese [21, 22], or Persian L1 [23] or some other variety of Arabic than PA [24].

We end with a word of caution. Our perceptual labelling method will be rather more complicated when also used to study dynamically changing vowel sounds such as diphthongs. Diphthongal trajectories can be synthesized with no problem but the number of stimuli will soon become unmanageable.

7. REFERENCES

[1] Ladefoged, P., Johnson, K. 2011. *A course in Phonetics*. Wadsworth.
[2] Ladefoged, P., Disner, S. 2012. *Vowels and consonants*. Wiley-Blackwell.
[3] Yavaş, M. 2011. *Applied English phonology*. Wiley-Blackwell.
[4] Thelwall, R., Sa'Adeddin, M. A. 1999. Illustrations of the IPA: Arabic. *Handbook of the IPA*. Cambridge University Press, 51–54.
[5] Palva, H. 1965. *Lower Galilean Arabic: An analysis of its anaptyctic and prothetic vowels with sample texts*. Studia Orientalia.
[6] Shahin, K. 2011. Palestinian Arabic. In: Versteegh, K. Eid, M., Elgibali, A., Woidich, M., Zaborski, A. (eds.), *Encyclopedia of Arabic language and linguistics, Vol.* 3. Brill, 526–538.
[7] Amir, N., Amir, O., Rosenhouse, J. 2014. Colloquial Arabic vowels in Israel: A comparative acoustic study of two dialects. *J. Acoust. Soc. Am.* 136, 1895–1907. doi: 10.1121/1.4894725
[8] Hall, N. 2017. Phonetic neutralization in Palestinian Arabic vowel shortening, with implications for lexical organization. *Glossa* 2, 1–23. doi: 10.5334/gjgl.257]
[9] Celce-Murcia, M., Brinton, D. M., Goodwin, J. M. Griner, B. 2010. *Teaching pronunciation: A course book and reference guide*. Cambridge University Press.
[10] Ladefoged, P. 1999. Illustrations of the IPA: American English. *Handbook of the IPA*. Cambridge University Press, 41–44.
[11] van Heuven, V. J., Afshar, N., Disner, S. 2020. Mapping perceptual vowel spaces in native and foreign language: Persian learners of English compared with American native speakers. In: Bátyi, S., Lengyel, Z. (eds.), *Bilingualism: Hungarian and non-Hungarian context. Studies in honor of Judit Navracsics*. Pannon Egyetem, 113–130. http://hdl.handle.net/1887/3146383
[12] van Heuven, V. J., Farran, B. M. (2022). The role of duration in the articulation and mental representation of American English vowels by Palestinian Arabic learners of English. 23rd Summer School of Psycholinguistics, Veszprém, www.researchgate.net/publication/361570474
[12] Afshar, N., van Heuven, V. J. 2022. Perceptual assimilation of English vowels by monolingual and bilingual learners in Iran. *Argumentum* 18, 172–191. doi: 10.34103/ARGUMENTUM/2022/9
[13] Boersma, P., Weenink, D. 2019. Praat, A system for doing phonetics by computer. www.praat.org
[14] Farran, B. M. 2022. The production and perception of American English sounds by Palestinian Arabic adolescents. PhD. diss., University of Pannonia. doi: 10.18136/PE.2022.820
[15] Hombert, J.-M., Puech, G. 1984. Espace vocalique et structuration perceptuelle: application au Swahili. *Pholia* 1, 199–208. http://www.ddl.cnrs.fr/Download/Pholia/Pholia_N-1.pdf
[16] van Zanten, E., van Heuven V. J. 1984. The Indonesian vowels as pronounced and perceived by Toba Batak, Sundanese and Javanese speakers. *Contr. Royal Inst. Anthrop*. 140, 497–521. doi: 10.1163/22134379-90003411
[17] Heeringa, W., Van de Velde, H. 2018. Visible Vowels: A tool for the visualization of vowel variation. *Proc. CLARIN 2018*, 120–123. https://office.clarin.eu/v/CE-2018-1292-CLARIN2018_Conference Proceedings.pdf
[18] Wang, H. 2007. *English as a lingua franca. Mutual intelligibility of American, Chinese and Dutch speakers of English*. LOT. https://www. lotpublications.nl/documents/147_fulltext.pdf
[19] Wang, H., van Heuven, V. J. 2006. Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In: van de Weijer, J. M., Los, B. (eds.), *Linguistics in the Netherlands 2006*. John Benjamins, 237–248. doi: 10.1075/avt.23.23wan
[20] Hillenbrand, J. M., Clark, M. J., Houde, R. A. 2000. Some effects of duration on vowel recognition. *J. Acoust. Soc. Am.* 108, 3013–3022. doi: 10.1121/1.1323463
[21] Perwitasari, A., Klamer, M., and Schiller, N. O. 2017. Quality of Javanese and Sundanese vowels. *J. Southeast Asian Ling. Soc.* 10, 1–9. http://hdl.handle.net/10524/52406.
[22] Perwitasari, A. 2018. *The acquisition of English vowels by Javanese and Sundanese native speakers*. LOT. https://www.lotpublications.nl/Documents/532_fulltext.pdf
[23] Afshar, N. 2022. Production and perceptual representation of American English vowel sounds by monolingual Persian and early bilingual Azerbaijani-Persian adolescents. PhD. diss., University of Pannonia. doi: 10.13140/RG.2.2.36272.02563
[24] Munro, M. J. 1993. Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Lang. Speech* 36, 39–66. doi: 10.1177/002383099303600103

[1] The authors are grateful to professor Sandra Ferrari Disner of the Linguistics Department at USC for making the materials and software to do the vowel identification experiment available to her students, running a quick survey of the participants' linguistic background and for collecting and transmitting the data files.
[2] We thank professor Hongyan Wang of Shenzhen University (P. R. China), who made her vowel data available to us. No results for the tokens of *hawed* are given in [18, 19] but the formant and duration values for the *hawed* tokens were included in the underlying dataset.
[3] Correlation of production and perceptual representation was considerably poorer at the level of individual speakers (for details and statistical tests, see [14])