The detailed comments Dr. Szilárd Szentgyörgy provided on reviewing my final dissertation are much appreciated. The following responses are available respectively:

Answer to the Chapter Reviews

> **Chapter 6** presents a contrastive acoustic analysis of vowels since it is critical that such cross-linguistic perceptual similarities be established in order to predict L2 learning difficulties more accurately.
>
> I am not sure this is a correct representation of what I tried to do. The contrastive acoustic analysis is not used to predict learning difficulties. Rather, the deviant configurations of vowel centroids and/or overlap between spreading ellipses test the results of learning difficulties rather than predict them.

Answers to the Comments on Chapters

Chapter 1

A. The reviewer is absolutely right. There will be differences in the exact pronunciation of the target vowels which depend on the local or regional variety of American English of the speaker. However, it is clear from my results that the variability between the speakers is relatively small. With the exception of the low back vowels, the remaining nine vowels have stable centroids and relatively small dispersion ellipses, which proves that the 20 speakers were in high agreement in their realization of the AE vowels, and – crucially – in much better agreement than the nonnative speakers. The large overlap between the two low back vowels can be expected and explained as the result of the low-back vowel merger, which affects many present-day regional varieties, including the Californian varieties.

B. Regarding the vowels listed as the focus of attention, this would be an example of where the reviewer's professional opinion is different from the one reflected by the dissertation. In the dissertation I have adopted the analysis of the AE vowel system advocated by Celce-Murcia et al. (2010). This is arguably the most authoritative text book of teaching the pronunciation of AE to foreigners. The authors divide the English vowels into simple vowels and complex vowels. The complex vowels are the true diphthongs /ai, au, ɔi/. All other vowels are simplex, and therefore treated as monophthongs. For a vowel to be classified as a diphthong, according to Celce-Murcia et al., the arrow in the vowel diagram that connects the starting point and terminal point of the diphthong MUST traverse

through an area taken up by some other vowel in the diagram. As long as the change from starting point to finish is between immediately adjacent vowel qualities, the vowel does NOT qualify as a diphthong. Other textbooks call vowels with such small quality changes VIRTUAL diphthongs. But, just as virtual reality is not reality, virtual diphthongs ae not really diphthongs. The most important reason why I (and others) include the vowels in *pain* and *bone* in the set of monophthongs is that these vowels remain perfectly identifiable if the small glide is omitted. When the glide is omitted from true diphthongs, they are no longer different from each other and/or some other vowels. For instance, the contrast between *find* and *found* will be lost, or between *pond* and *pound* or between *tall* and *toil*.

C. Regarding the tenseness-laxness in a phonetic sense, it is not that simple. Duration covaries with tenseness. Because the articulators have to assume more peripheral configurations it takes more time to reach these positions. Therefore, tense vowels are also longer vowels.

D. I am familiar with these observations. In fact, I explicitly mention this on p. 90 of the dissertation, where I write: “In the control data /æ/, although phonologically lax on distributional grounds, …”

E. Regarding the problem arises in connection with vowel length, I beg to differ here. It may be a consequence of the neutralization but it is certainly NOT a necessary consequence. The fact that a cue is neutralized in one particular condition does not mean that the cue would not work in contexts where there is no neutralization. For instance, German, Dutch, and in fact most languages in the world, neutralize the voiced-voiceless contrast in word-final position (English is among the minority of languages that do not have final devoicing of obstruents. But Germans and Dutchman are perfectly able to hear and produce the contrast in word-initial and medial position. But I certainly admit that it would be important to co-vary the voiced-voiceless nature (as well as the stop-affricate-fricative nature of the final consonant) AND the vowel duration AND the vowel quality in future experiments. However, co-varying all these factors would yield a very large experiment, on a scale that could not be incorporated in the tight schedule Pannon Egyetem puts on writing doctoral dissertations. Consequently, I would call such additional experiments an challenge for the immediate future rather than an opportunity missed.

Chapter 2

A. In regards to the problems concerning the treatment of vowel length, this assumes that /æ/ is short (or as short as other short vowels) before a voiceless consonant. The literature on AE shows that this is NOT the case. The vowel in *cass*, *cat*, *calf*, *cap*, etc. is munch longer than the vowel in *kiss*, *kit*, *cliff*, *tip*, etc. I refer here to House (1961), who presents the vowel durations of American English, broken down by the contexts mentioned by the reviewer. So, the conclusion must be that /æ/ is unconditionally a long vowel in terms of its duration (but not in terms of its phonotactic distribution).
B. Concerning the basis for treating central or centralized vowels differently in English and Azerbaijani, in Persian and Azerbaijani there is no difference in vowel duration – therefore the idea of a tense-lax contrast does not apply.

Chapter 4

A. Again, vowel duration is not contrastive in Persian and Azerbaijani. I am only willing to consider a difference in vowel quality as a cue in a tense-lax contrast if it co-varies with duration.
B. I do not think that the monophthongal pronunciation of /e/ by EFL learners will make this vowel indistinct from short /ɪ/ or /ɛ/. so. There will still be a large difference in vowel duration that will keep short raised /ɛ/ separate from long monophthongal /e/.

   This is also supported by the fact that even non-native speakers tend to identify /e/ as an /eɪ/ diphthong. The part highlighted in green has the relationship reversed: my Persian participants do not identify their own /e/ as AE /eɪ/. Instead, they identify AE /eɪ/ as a non-typical exemplar of their own /e/-vowel. In spite of the audible diphthongization, the foreign vowel is still classified as a token of Persian /e/ (i.e., a monophthong). As a result of this pairing (or: assimilation), we expect the Persian learner of AE to substitute their monophthongal /e/ for the AE semi-diphthong [eɪ], at least in the early stages of AE acquisition. Note also that AE [eɪ] is both too long and too diphthongal to be a close match to Persian /e/. My experiment does not allow us to choose between these two reasons. Since listeners are generally more sensitive to mismatches in duration than to mismatches in vowel quality, the long duration of AE [eɪ] may well be the primary reason for the judged non-typicality.

Chapter 5

A. In regards to the comment on the variety of American English, I can say that all this is true, but one has to be practical. It is very difficult (and costly) to find large numbers of American listeners who belong to homogeneous language communities. In fact, I consider myself lucky that we found a colleague willing to run the experiment for us in Los Angeles, with a majority of listeners speaking roughly the same regional variety. It is true that the /ɑ, ɔ/ categories proved indistinct in the responses (suggesting indeed the effect of the low-back merger), but for other vowel categories we found clearly different centroids and non-overlapping spreading ellipses. This, to me, suggests that the 20 respondents are sufficiently homogeneous and representative for General American English. Since my study was intended to identify learning problems for nonnative speakers of American English, I can practically ignore the effect of the low-back merger, because native speakers are no longer sensitive to the contrast. Celce-Murcia et al. (2010) and many other textbooks give the advice to EFL teachers not to spend any time on the /ɑ, ɔ/ contrast.

B. About the potential problem with the keywords ending in /l/, we corresponded about this issue before. I do not believe that the velarization caused by dark /l/ influences the mental conception (perceptual representation) of the vowel that is exemplified by the keyword. My results seem to corroborate this assumption. Also, there is circumstantial evidence that the assumption is correct. In so-called phoneme detection experiments, the participant is told to press a key as quickly as possible as soon as they hear a pregiven target sound, e.g., the sound /p/ as in "pain". Results of such experiments bear out that the word used to identify the target is irrelevant to the execution of the detection task. But the reviewer's point is well taken. In the journal article we are currently preparing, we will address this issue so as to pre-empt the potential objection to our instructions in the vowel identification task.

Chapter 6

A. I thank the reviewer's comment on "Incorrect pronunciations occurred in the vast majority of the tokens of *sawed* and *hawed*, which were then pronounced with a full diphthong /ɑʊ/ (as in *cloud*)." This explanation makes eminent sense.

October 27st, 2022, Veszprém

Naeimeh Afshar

University of Pannonia