FORMANT CHARACTERISTICS OF UNSTRESSED UNREDUCED VOWELS IN AMERICAN ENGLISH: ONLY EXPLAINED BY DURATION?

Mariko Sugahara

Doshisha University, Japan
msugahar@mail.doshisha.ac.jp

ABSTRACT

It has been considered that there are four levels in English vowels: primarily stressed (P), secondarily stressed (S), so-called "unstressed unreduced" (U) and unstressed reduced (R). According to Fear et al., U vowels were shorter than P and S but longer than R, and they were more reduced than P and S though not as reduced as R in terms of their vowel qualities. The question asked in this study is whether U's partial reduction is only a by-product of its shorter duration. A production experiment was conducted with 25 American English speakers, focusing on /æ/ and /ɔː~ɑː/. We performed statistical analyses using LMM and the results indicate for both of the vowels that U's partial phonetic reduction is not simply a by-product of its shorter duration.

Keywords: unstressed unreduced vowels, American English, phonetic reduction, duration, formant characteristics

1. INTRODUCTION

It has been considered that there are four levels in English vowels: primary stress (P), secondary stress (S), so-called "unstressed unreduced" vowels (U), and unstressed reduced vowels known as schwa ([1], [2]). According to the production study conducted by [1], the acoustic characteristics of U were in between those of the stressed vowels (P and S) and R. U vowels were shorter than P and S but longer than R, and they were more reduced than P and S though not as reduced as R in terms of their vowel qualities which are indicated by *F2-F1* differences. Nonetheless, the perception experiment conducted by [1], in which the four types of vowels were cross-spliced and listeners rated the acceptability of the resulting words, the listeners more steadily grouped U with the stressed vowels (P and S) rather than R. This suggests that the U vowels are perceptually categorized as full vowels just as P and S are despite their partial spectral reduction.

According to [3], [4] and [5], the major source of vowels' phonetic reduction in their spectral characteristics is their shorter duration: The shorter the duration of a vowel is, the more difficult for its articulators to reach their targets, resulting in reduction in their spectral characteristics. This is called target undershoot. The strongest hypothesis under this undershoot model is that English U vowels are primarily associated with shorter duration than P and S, and the reduction of their spectral characteristics is nothing but a by-product of their shorter duration.

However, it is still possible to imagine that the articulatory target of U vowels is set at a slightly slacker position than that of P and S vowels independently of U's shorter duration, and the spectral characteristics of U may differ from those of P and S even when the effect of duration is removed.

A production experiment was conducted and statistical analyses using LMM (Linear Mixed Model) were performed to test this alternative view.

2. PRODUCTION EXPERIMENT

2.1. Participants

The participants were 25 paid native speakers of American English, comprising 15 females and 10 males, who were undergraduate or graduate students at universities in the United States, from 18 to 29 years old at the time of the experiment. Eight of them were from the New England variety area, six from the West, five from the South, three from the Midland, one from the North Central area, one from the Inland North area. One person was raised in multiple areas.

2.2. Word sets

Among all the word sets used in our production experiment, 16 words were considered in the subsequent analyses, which are shown in Sets 1-4. They all started with non-high monophthongs. The initial vowels of the four words in each set were divided into the four types: P, S, U, and R. The words in Sets 3 & 4 except for *authenticity* and *Athena* were adopted from [1].

/æ/ as the initial full vowel (P, S, U)

Set 1	(P) ˈ**a**ctive	(U) **a**cˈtivity
	(S) ˌ**a**ctiˈvation	(R) **a**/ə/cˈcept
Set 2	(P) ˈ**a**sphalt	(U) **a**sˈphaltic
	(S) ˌ**a**spiˈration	(R) **a**/ə/sˈparagus

/ɔː~ɑː/ as the initial full vowel (P, S U)

Set 3 (P) ˈ**au**diences (U) **au**ˈdition
(S) ˌ**au**diˈtoria (R) **a**/ə/ˈddition

Set 4 (P) ˈ**au**thorize (U) **au**ˈthentic
(S) ˌ**au**thenˈticity (R) **A**/ə/ˈthena

The initial vowels of P, S and U in Sets 1 &2 were /æ/, and those in Sets 3 & 4 were /ɔː~ɑː/. Whether they are pronounced as /ɔː/ or /ɑː/ depends on speakers' varieties ([6], [7], [8]). (The spectral variation of /ɔː~ɑː/, however, seems to be even smaller than that of /æ/, as shown in Figures 2 and 3).

The initial vowels of all the four words in each set were followed by the same consonant so that the effect of the following consonant is kept the same across the members in the same set. We included the reduced vowel /ə/ as the control case in order to verify if the U vowels keep their unreduced vowel qualities.

Additional word sets were also included in the production experiment, which were not considered in this study. It is because the initial vowels in the additional sets did not satisfy the criteria of being a non-high monophthong with no onset consonant.

2.3. Recordings

The recordings took place in a sound-proof studio. Cardioid condenser microphones are used: Countryman Headset Microphone for 17 participants and RØDE NT2-A for eight participants. Sound was recorded as WAV files in monaural using Marantz Solid State Recorder at 16 bits with the sampling rate of 44.1 kHz.

The target words were embedded in two types of carrier sentences: (i) *What did you say? I said "___" this time*, and (ii) *I didn't say "___." X said "___."* For each word, the carrier sentences (i) and (ii) were produced once respectively. About 300 filler sentences were also included in each speaker's recording session and the presentation order of sentences was randomized. This study, however, only focused on the target word tokens produced in the carrier sentences of type (i), where the word was interpreted as new and its primary-stress syllable coincided with a nuclear pitch accent.

2.4. Acoustic measurements

Acoustic data were obtained from the initial vowels of the 16 target words in Sets 1 to 4. Although the total number of the initial vowels should be 400 (4 words*4 sets*25 speakers), four tokens were missing (two tokens of *accept* and two tokens of *authorize*) and the number of recorded tokens considered in this study was 396 (16 from 21 participants and 15 from four participants).

The WAV file of each token was segmented using PRAAT ([9]), and we obtained (i) the duration values of the initial vowel intervals in ms, and (ii) their formant values (*F1, F2*) in Hz at their midpoint.

The starting point of each vowel was defined as where its formants explicitly appeared. The point sometimes coincided with the beginning of glottalized phonation. Occasionally, there were cases in which an epenthetic schwa vowel followed the coda consonant /d/ of the preceding word *said* and no clear break in the spectrogram was available between the epenthetic vowel and the beginning of the target initial vowel. In such cases, the starting point was identified as where the amplitude of the waveform narrowed down or where the vowel quality of the epenthetic vowel changed to that of the target initial vowel. The ending point of each vowel was defined as where consonant frication or closure started.

The *F1* and *F2* values in Hz were converted into the Bark scale, and the durational values were converted into the logarithmic scale (Ln). We regard the distance between the Bark-scaled *F1* and the Bark-scaled *F2*, i.e., *F2-F1* henceforth, as the indicator of vowel reduction following [1] that greater *F2-F1* indicates more reduction. We adopted this measure for empirical reasons. Our speakers' *F1* of /æ/ decreased while that of /ɔː~ɑː/ was almost unchanged, and their *F2* of /ɔː~ɑː/ increased while that of /æ/ was almost intact, as the stress level shifted from P to U. In order to analyse the spectral characteristics of the two vowels in a parallel manner, the measure of *F2-F1* was needed.

The outliers of the *F2-F1* values were identified separately for each stress level of each vowel using the 1.5*IQR rule. There were ten *F2-F1* outliers: one instance of *activation*, *active* and *authorize* each, two instances of *audition* and *authentic* each, three instances of *authenticity*, which were removed when creating Fig. 1 and Fig. 3, and when conducting statistical analyses. After the removal of the ten outliers, the number of the *F2-F1* data was 386: P=96, S=100, U=98, R=92.

3. PRELIMINARY OBSERVATIONS

Before performing the main statistical analyses, we made some preliminary observations.

3.1. Duration and *F2-F1*

Given that a greater *F2-F1* value is an indicator of phonetic reduction, and if shorter duration incurs more phonetic reduction, we expect *F2-F1* to increase as vowel's duration decreases.

To see the relationship of duration and *F2-F1*, durational values were normalized to remove speaker differences. First, each vowel token's duration was converted into the logarithmic (Ln) scale. Secondly, for each speaker, the mean of the log durations was obtained separately from the initial vowels of the eight words in Sets 1 & 2 (i.e., the /æ/ sets) and from those in Sets 3 & 4 (i.e., the /ɔː~ɑː/ sets). For example, for each speaker, the mean duration of the initial vowels of P, S, U, R in Sets 1 & 2 and the mean duration of those in Sets 3 & 4 were obtained separately. Thirdly, we subtracted the mean duration from the duration of the corresponding vowel token, which is the normalized duration value.

The box plots in Fig 1. summarize the normalized duration of the four vowel levels, and the scatter plots in Fig. 2 show the normalized duration values plotted on the x-axis and *F2-F1* plotted on the y-axis. It is evident from Fig. 1 that as the stress level decreased from P to R, duration also decreased, which replicates the observation by [1].

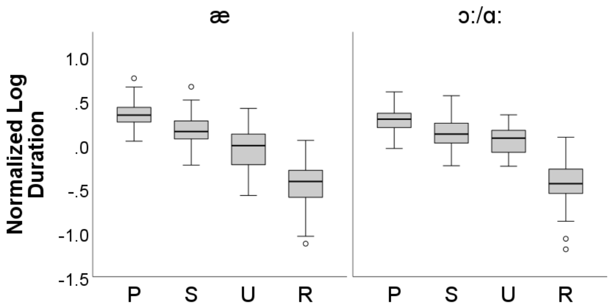

Figure 1: Normalized duration of the four levels

Fig. 2 shows a negative correlation between the *F2-F1* values and the duration values across the four different vowel levels. There are, however, differences between U and the other vowel types: U's slopes of fitted lines are steeper than the other groups' slopes. We test if the differences are significant in the following statistical analyses.

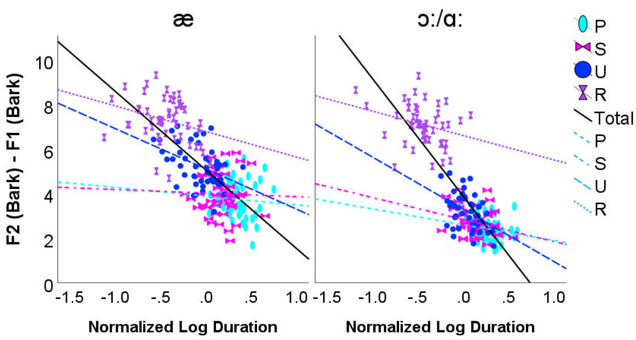

Figure 2: *F2-F1* plotted against normalized duration

Another observation from Fig. 2 is that the overlap between U's and R's data points for /æ/ is greater than those for /ɔː~ɑː/ on both the x-coordinate (duration) and the y-coordinate (*F2-F1*). This implies that the degree of U's phonetic reduction is different between /æ/ and /ɔː~ɑː/. This point will be discussed in Sec. 5.

3.3. Sex and *F2-F1*

Since *F1* and *F2* values are known to vary according to vocal tract length ([10] among others), sex may affect the values of *F2-F1*. Fig. 3 presents the box plots of *F2-F1* from females and males. Although the ranges of the boxes and the whiskers slightly differed between F and M, their medians were approximately the same for all of the vowel types regardless of the vowel differences. It could be argued that normalisation between the two sex groups was achieved to some extent by taking the difference between *F2* and *F1*.

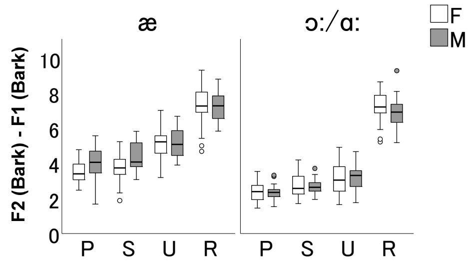

Figure 3: *F2-F1* box plots of females and males

4. STATISTICAL ANALYSES AND RESULTS

4.1. Analyses

Statistical analyses were performed on SPSS using LMM (Linear Mixed Model) to investigate whether (i) the vowel types and duration both contribute to predicting *F2-F1*, (ii) *F2-F1* differs between U and the other vowel types even when given the same duration, and (iii) the relationship between *F2-F1* and duration is the same across the four vowel types.

For each of the vowels, three models were tested: M1, M2 and M3. The target variable was *F2-F1* in all of the models. One fixed factor (*Type*, i.e., P, S, U, R) was included in M1, two factors (*Type* and *Duration*) in M2, and three factors (*Type, Duration, Type*Duration*) in M3. We consider collinearity between *Type* and *Duration* not to be a problem here because their VIF was 2.94 for /æ/ and 2.69 for /ɔː~ɑː/, less than 5. (The VIF scores were obtained by converting *Type* into a scalar variable.) Random intercepts of *Speaker* and those of *WordSet* were also included in the three models.

4.2. Results

First, the three models were compared. For both of the vowels, M3's AIC was the smallest (for /æ/, 501.76 for M1, 489.81 for M2, 481.46 for M3; for /ɔː~ɑː/, 404.79 for M1, 381.61 for M2, 374.79 for M3). From this, M3 seems to be the most appropriate model among the three. However, not all of the three fixed factors in M3 were significant. Although M3 itself was significant: [*F* (7, 188) =

102.9, $p < .001$ for /æ/; $F(7, 182) = 308.3$, $p < .001$ for /ɔː~ɑː/], and *Type* and *Duration* in M3 were also significant respectively: [*Type*: $F(3, 188) = 43.65$, $p < .001$ for /æ/ and $F(3, 182) = 112.06$, $p < .001$ for /ɔː~ɑː/; *Duration*: $F(1, 188) = 6.77$, $p < .025$ for /æ/ and $F(1, 182) = 22.48$, $p < .001$ for /ɔː~ɑː/], the factor of *Type***Duration* was not significant for both of the vowels: [$F(3, 188) = 1.56$, $p = .2$ for /æ/; $F(3, 182) = 1.12$, $p = .34$ for /ɔː~ɑː/]. (Since there are separate analyses for the two, we adopted Bonferroni correction and the α level was set at .025.) Tables 1 and 2 summarize the results of the coefficients of the fixed factors in M3.

Table 1: Results of the fixed factors for /æ/ (M3)

	β	SE	t	95%CI	
Intercept (U)	5.1	.20	25.9***	4.7	5.5
diff (P-U)	-1.2	.29	-4.2***	-1.8	-.66
(S-U)	-1.1	.19	-6.0***	-1.5	-.77
(R-U)	1.7	.25	6.7***	1.2	2.2
Slope (U*D)	-1.6	.51	-3.2**	-2.6	-.60
diff (P*D-U*D)	1.2	.90	1.3	-.59	3.0
(S*D-U*D)	1.7	.85	2.0	.05	3.4
(R*D-U*D)	.49	.69	.71	-.87	1.8

'***' <.001, '**' <.01, '*' <.025

Table 2: Results of the fixed factors for /ɔː~ɑː/ (M3)

	β	SE	t	95%CI	
Intercept (U)	3.3	.19	17.5***	2.9	3.7
diff (P-U)	-.57	.23	-2.5*	-1.0	-.12
(S-U)	-.51	.14	-3.5***	-.79	-.22
(R-U)	3.2	.21	15.8***	2.9	3.7
Slope (U*D)	-2.2	.60	-3.6***	-3.4	-.98
diff (P*D-U*D)	.99	.91	1.1	-.81	2.8
(S*D-U*D)	1.5	.84	1.8	-.16	3.2
(R*D-U*D)	1.0	.73	1.4	-.42	2.5

The tables above show that for both of the vowel phonemes, U's estimated slope and intercept were significant. The term 'diff (*TypeX*-U)' represents the estimated difference between U's intercept and the intercepts of the other vowel types' slopes. The negative values of 'diff (P-U)' and 'diff (S-U)' mean that U's intercept is significantly greater than P's and S's whereas the positive value of 'diff (R-U)' means that U's intercept is significantly smaller than R's.

The estimated differences in slopes between U and the other vowel types represented as 'diff (*TypeX**D-U*D)' were not significant, which is already expected as mentioned above. (Because U's slope was negative, their positive values mean less steep slopes.)

5. DISCUSSION AND CONCLUSION

The results above show that *Vowel Type* contributes to predicting *F2-F1* together with *Duration*. U's intercept being greater than P/S's and it being smaller than R's means that when given zero on the x-coordinate, i.e., the coordinate of *Duration*, U's estimated *F2-F1* value is expected to be greater than P/S's and smaller than R's. The value of zero on the x-coordinate corresponds to the mean of the normalized durations in our study. As already mentioned in Sec. 3.1, we normalized durations by taking deviations from the mean of the initial vowels' durations for each vowel phoneme group. Because they are deviations, their mean is zero. Given this, U's estimated *F2-F1* is greater than P/S's and smaller than R's when given the durational mean of all the tokens. This suggests that U's spectral quality is more reduced than P's and S's, and less reduced than R's independently of duration, i.e., U's partial reduction is not simply a by-product of its shorter duration. This further implies that U's articulatory target is set at a slightly slacker position than P's and S's.

Although U's slopes of fitted lines were steeper than the other vowel types' slopes (see Fig. 2), the difference was not statistically significant. That is, there is little persuasive evidence to support the claim that the U's relation with duration is different from the other vowels' relation to it.

One thing to note is that the magnitude of U's phonetic reduction is more pronounced for /æ/ than for /ɔː~ɑː/ as mentioned in Sec. 3.2. A key for explaining this may be that the former is lax while the latter is tense. Lax vowels have been claimed to be more prone to reduction than tense vowels in metrically weak position ([11]), with which our outcome seems to comply. Because the only lax and tense vowels tested here are /æ/ and /ɔː~ɑː/ respectively, it is too early to draw any firm conclusion regarding this, which needs to be investigated further.

Related to this point, a perception study may be necessary, too. In their perception study, [1] found that U was grouped with P and S, i.e., the full vowels, rather than R by native speakers. They, however, only used tense vowels. Given a possibility that the lax vowel /æ/ of the U type is more reduced than the tense vowel /ɔː~ɑː/ of the same type, it is still unresolved how the former is perceptually grouped.

The target vowels considered in this work were limited to non-high monophthongs preceded by no onset consonant. We need to look at other vowels, too. The presence or absence of onset consonants may also affect the outcome. As mentioned in Sec. 2.2, additional word sets with other vowels with and without onset consonants were also recorded in our production experiment. Their spectral patterns will be considered in our future study. We have also recorded the target words in a post-focus environment as mentioned in Sec. 3.3., and it is of interest to us how U vowels behave in that environment.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP17K02828 and JP22K00627. We are thankful to reviewers for their valuable comments, which helped us improve the quality of the paper. All errors are my own.

REFERENCES

[1] Fear, B. D., Cutler, A., Butterfield, S. 1995. The strong/weak syllable distinction in English. *J. Acoust. Soc. Am*. 97, 1893-1904.
[2] Ladefoged, P., Johnson, K. 2015. *A Course in Phonetics*, 7th ed. Cengage Learning.
[3] Lindblom, B. 1963. Spectrographic Study of Vowel Reduction. *J. Acoust. Soc. Am*. 35, 1773-1781.
[4] Moon S-J., Lindblom, S. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *J. Acoust. Soc. Am*. 96, 40-55.
[5] Barns, J. 2006. *Strength and Weakness at the Interface.* Mouton de Gruyter.
[6] Jones, D., Roach, P., Setter, J., Esling, J. 2011. *Cambridge English Pronouncing Dictionary*, 18th ed. Cambridge University Press.
[7] Upton, C., & Kretzschmar, Jr. W. A. 2017. *The Routledge Dictionary of Pronunciation for Current English*, 2nd ed. Routledge.
[8] Wells, J. C. 2008. *Longman Pronunciation Dictionary*, 3rd ed. Longman.
[9] Boersma, P., Weenink, D. 2022. *Praat: Doing Phonetics by Computer* [Computer program]. Version 6.1.51.
[10] Johnson, K. 2020. The ΔF method of vocal tract length normalization for vowels. *Laboratory Phonology*. 11(1), 1-16. DOI:https://doi.org/10.5334/labphon.196
[11] Chomsky, N., Halle, M. 1968. *The Sound Pattern of English*. MIT Press.