REGIONAL AND INDIVIDUAL VARIATION IN ACOUSTIC TARGETS OF /AI/ and /AU/ IN AMERICAN ENGLISH

Irina Shport[1], Marie Bissell[2], Kelly Berkson[3], Katie Carmichael[4]

[1]Department of English, Louisiana State University; [2]Department of Linguistics, The Ohio State University; [3]Department of Linguistics, Indiana University; [4]Department of English, Virginia Tech, United States
ishport@lsu.edu, bissell.43@osu.edu, kberkson@iu.edu, katcarm@vt.edu

ABSTRACT

English diphthongs are represented as bisegmental (e.g., /aɪ/) and assumed to have two acoustic targets (onset, offset). These phones are standardly represented with variable symbology (e.g., /ai/, /aj/), and previous work indeed reports variability in diphthong offsets. To investigate whether variation can be explained by dialectal and/or individual differences, we examined diphthongs (/aɪ, aʊ/) produced by 41 speakers of American English from Ohio (Midland, Northern regions) and Louisiana (Southern region).

Comparison of offset spectral estimates to nearby monophthongs indicate that the /aɪ/ offset was relatively consistently acoustically close to [ɪ], but the /aʊ/ offset was highly variable for both groups. While some of these findings, such as the degree of dynamic spectral change in diphthongs, can be explained by dialectal differences, it is also possible that the diphthongs have different underlying structure (e.g., more clearly biphasic onset-offset for /aɪ/ than for /aʊ/).

Keywords: diphthongs, nearby monophthongs, positional relationship, cross-dialectal differences.

1. INTRODUCTION

This study examined spectral trajectories of two American English diphthongs, /aɪ/ and /aʊ/. In speech sciences, three sets of notations are used to represent these phones. The offglide targets are variably represented as glides, lax high vowels, or tense high vowels [1]. These three sets of notations – /aj/-/aw/, /aɪ/-/aʊ/, /ai/-/au/ – bear inherent theoretical assumptions about the underlying targets of diphthong offglides, which are rarely critically examined in context of variation in vowels systems across English dialects.

Herein, we investigate whether phonological representation of the offsets as high vowels is (a) accurate for both diphthongs and (b) supported by robust acoustic data in word productions by English native speakers from two dialectal regions. Diphthongs are often assumed to have two steady-state targets as their onset and offset [2]. Yet the single study (to our knowledge) that has investigated what acoustic properties of English diphthongs are most relevant to their classification tested three models – onset plus offset, onset plus slope, onset plus direction – and yielded correct classifications of a diphthong corpus exceeding 90% accuracy [3]. Further, a study comparing the English and German equivalents of these two diphthongs suggested that they may have different onsets in English, concluding that /a/ accurately represents the onset of /aɪ/ but not /aʊ/ [4]. Both these works, though very different, raise questions about how best to represent the diphthongs.

The use of the IPA phonetic symbols for lax vowels in English diphthongs may be traced back to Daniel Jones (1881-1967) [1, footnote 1]. The /aɪ/ and /aʊ/ notations took deep root in the linguistic community and are commonly used by students in phonetics classes and by researchers alike (also, in this paper). Research findings, however, have showed that the acoustic targets of the diphthongs can approximate a variety of phones for both onset (/æ/, /a/, /ɑ/, or /ʌ/) and offset (/i/, /ɪ/, /ɛ/, /u/, /ɔ/, or /ʊ/), although they are more consistent for the former than the latter [2-3, 5-7]. For this reason, we chose to focus on offsets and, specifically, on the positional relationship between the estimates of offset targets and nearby monophthongs.

In sociolinguistic tradition, ARPABET phonetic symbols AY and AW are often used for vowels in *bite* and *bout* [8], representing the offset targets as underlying glides. Recent work suggests that while high vowels /i, u/ and their glide counterparts /j, w/ may belong to distinct phonemic categories, categorical distinction in their phonetic properties is not so much in the degree of constriction (as reflected in formant values), but in their temporal organization (as reflected in relative timing of articulatory gestures) [9]. Based on these findings and our focus, we excluded the /aj/ and /aw/ representations from consideration here.

In research so far, data were often collected in Midwestern states and might have been taken to represent General American English [3, 7]. Following [6], we contribute cross-dialectal data collected in a variety of phonetic contexts to this line of research. Because the phonological voicing status of the

following coda was shown to significantly alter the duration and spectral properties of vowels (e.g., American English in [10-11]; Australian English in [11]; Canadian raising in [13-15]) we examined formant values of diphthongs when followed by voiced and voiceless segments. Because speech in Southern U.S. communities is often characterized by /aɪ/-monophthongization and resistance to back vowel fronting which would affect diphthong offglides, participants were recruited in the state of Louisiana to represent the Southern region and in the state of Ohio to represent the Midland and Northern regions [16]. In what follows, we examine whether offglides approximate tense or lax high vowels (or other vowels), and whether speaker dialect or voicing context can explain offset variability in the target diphthongs.

2. METHODS

2.1. Data collection

Data was collected as a part of a larger study in which the two diphthongs were elicited in pre-voiced and pre-voiceless contexts (Table 1). Words with /aɪ/ totalled 52 and words with /aʊ/ totalled 35. In addition, thirteen monophthong vowels /iː, ɪ, eː, ɛ, æ, ə, ɜː, ɑ, ɔ, ʌ, oː, ʊ, u/ were elicited in b_t and h_d English word contexts (e.g., *beat/heed*, *bat/had*).

Target	Pre-Voiced Context	Pre-Voiceless Context
/aɪ/	*prize, cyber, rider, siding, vibration*	*price, viper, writer, biting, citation*
/aʊ/	*browse, cows, cloudy, however*	*house, couches, pouting, outlandish*

Table 1: Example stimulus words.

Participants were recruited in Columbus, Ohio (18 speakers, 14 females) and Baton Rouge, Louisiana (23 speakers, 15 females). They recorded themselves reading stimulus words one at a time in an online Qualtrics survey (words were randomized and presented three times each) and emailed the recordings to the investigators.

2.2. Analyses

All data was force-aligned using the MFA aligner [17]; then, segmentation was manually hand-checked in Praat [18]. A Praat script was used to extract first and second formant (F1, F2) values at 21 equidistant point from the onset of each vowel token, at 5%-increment into vowel duration. Next, we describe a set of analyses conducted for each speaker's data.

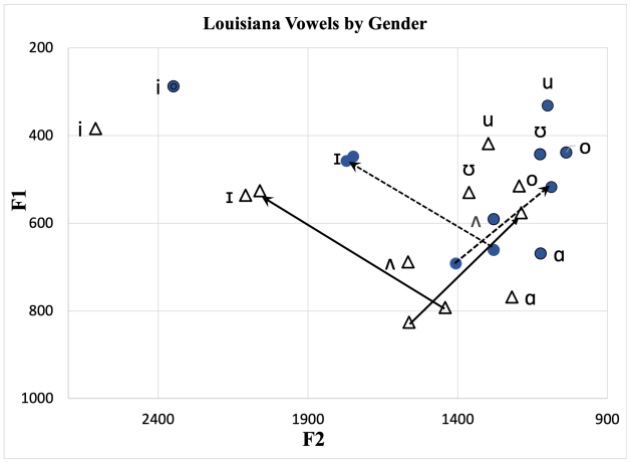
To provide a single estimate for each monophthong vowel (Mono), means of median values across 17 time points (i.e., 90% of vowel duration to exclude coarticulatory effects) were calculated [12]. To provide an estimate for the onset and offset of each diphthong (Di.on and Di.off), means of median values across the 3-5th and 17-19th time points (i.e., the 10-20% and 80-90% of vowel duration) were calculated [3]. After initial observations of individual F1/F2 plots, ten Euclidean distances (ED) between Mono and Di.off were calculated per speaker to estimate the distance between the /aɪ/'s offset and /i, ɪ, ʌ, æ, ɑ/ and between the /aʊ/'s offset and /u, o, ʌ, ʊ, ɑ/ (ED types). The formula in (1) was used to calculate EDs and served as a speaker-intrinsic normalization technique.

(1) $$ED_{/Mono/-/Di.off/} = \sqrt{(F1_{/Mono/} - F1_{/Di.off/})^2 + (F2_{/Mono/} - F2_{/Di.off/})^2}.$$

Mixed modeling was used for statistical analyses of these ED values, five ED types per diphthong.[1]

3. RESULTS

3.1. Variability in positional relationships



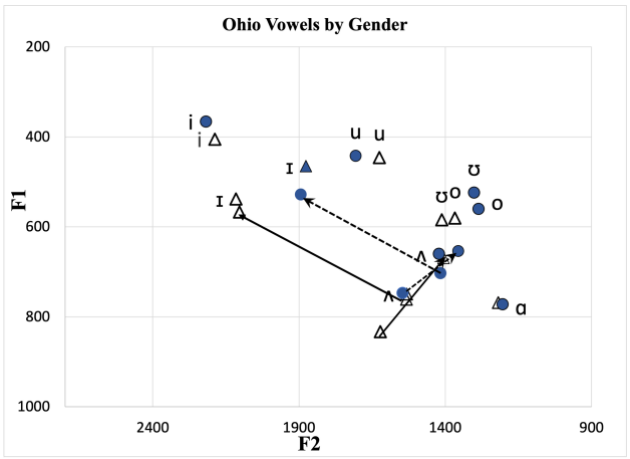


Figure 1: Mean vowel characteristics for two regions (females: open triangles, solid lines; males: filled circles, dashed lines). Monophthongs are represented by singular F1/F2 datapoints; diphthongs are represented by lines connecting the Di.on and Di.off estimates.

Figure 1 illustrates some regional differences in the vowel systems, such as [u] and [o] fronting (a relatively high F2) and [ʊ] and [o] lowering (a relatively high F1) in Ohio as compared to Louisiana. The degree of spectral change in diphthongs appears to be smaller in Ohio than Louisiana, especially for /aʊ/. In both plots, the offset of /aɪ/ is approximating [ɪ] the most. The offset target of /aʊ/ is less apparent: likely, [o ʊ] for Louisiana speakers, and [o ʊ ʌ] for Ohio speakers.

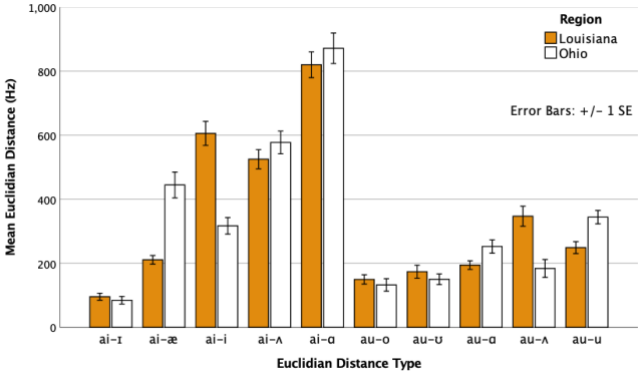


Figure 2: Summary of offglide positional relationship to nearby monophthongs (ten ED types) by region. Small ED values indicate relative proximity of diphthong offsets to monophthongs.

Generalized linear mixed modeling (GLMM) on ED values with region (2) and nearby monophthongs (5) as fixed factors showed a significant interaction between region and nearby monophthong for /aɪ/, $F(4,195) = 25.11$, $p < .001$, and for /aʊ/, $F(4,195) = 13.32$, $p < .001$. Figure 2 illustrates this interaction. For /aɪ/, the largest regional difference is in the offset proximity to [i] and [æ] ($t = 5.60$, $p < 0.001$, $t = 6.05$, $p < 0.001$). For /aʊ/, the largest regional difference is in the offset proximity to [u] and [ʌ] ($t = 3.39$, $p = 0.002$, $t = 3.79$, $p < 0.001$).

To examine individual variability in offset category membership, the numbers of speakers exhibiting specific primary and secondary proximal positions (i.e., the smallest and the second smallest EDs per diphthong) are summarized in Table 2. This summary confirms that for /aɪ/, the primary offset target is robust – the lax /ɪ/, with secondary proximal positions varied by region. For /aʊ/, the primary and secondary proximal positions largely varied.

Trends	LA	OH
aɪ - ɪ (i)	-	14 speakers
aɪ - ɪ (æ)	22 speakers	4 speakers
aɪ - æ (ɪ)	1 speaker	-
aɪ - ɪ (ʌ)	1 speaker	-
aʊ - o	12 speakers	7 speakers
aʊ - ʊ	7 speakers	2 speakers
aʊ - ʌ	-	8 speakers
aʊ - ɑ	4 speakers	-
aʊ - u	-	1 speaker

Table 2: A summary of speakers by absolute proximity of their diphthongs' offsets to selected monophthongs (i.e., ED types). The smallest ED was taken as the closest match. The second closest matches are in parentheses for the /aɪ/ offset but not for /aʊ/ because their number exceeded ten.

3.2. Voicing context

We also examined whether voicing of the following segment influences the acoustic realization of diphthong offglides, specifically, their F2 values at onset and offset.

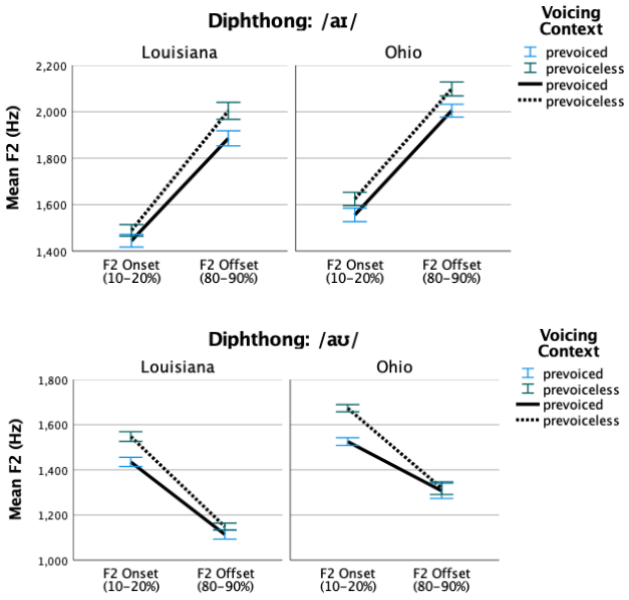


Figure 3: F2 estimates at diphthong onsets and offsets by region and the following voicing context.

GLLM analyses of F2 estimates with target (onset, offset), region (LA, OH), and voicing (pre-voiced, pre-voiceless) as fixed factors yielded a significant effect of voicing (/aɪ/: $F(1,322) = 32.83$, $p < .001$; /aʊ/: $F(1,322) = 46.07$, $p < .001$) and region (/aɪ/: $F(1,322) = 5.70$, $p = .018$; /aʊ/: $F(1,322) = 19.38$, $p < .001$). The interactions between voicing and region were not significant in either diphthong model. The interaction between voicing and target was significant for /aʊ/ only, $F(1,322) = 22.38$, $p < .001$. Figure 3 illustrates these results: F2 was higher in pre-

voiceless than pre-voiced contexts except for /aʊ/ offsets, and in productions of Ohioans than Louisianans.

4. DISCUSSION

This study addressed the question of whether acoustic data allow us to infer robust phonological offglide targets in productions of two diphthongs, and if so, what those targets are. We found that across two dialectal regions of North American English, the offset of /aɪ/ consistently approximated the monophthongal /ɪ/, although the secondary trend was for this offset to approximate [i] for Ohioans and [æ] for Louisianans. The latter secondary trend may be explained by /aɪ/ monophthongization, a still frequently observed characteristic of Southern speech where /aɪ/ does not have much of an offglide, with relatively small dynamic spectral change across the diphthong, as illustrated in Figure 4. The productions of /aɪ/ were overall more fronted by speakers from Ohio than speakers from Louisiana, and more in pre-voiceless than pre-voiced contexts. It is possible that if /aɪ/ offsets were examined only in a dialect such as Midland / Northern, in words where it is followed by a voiceless segment, a researcher may conclude that the offset target is closer to the tense /i/ rather than the lax /ɪ/. Based on our data, it is reasonable to infer that the underlying form of the diphthong is /aɪ/.

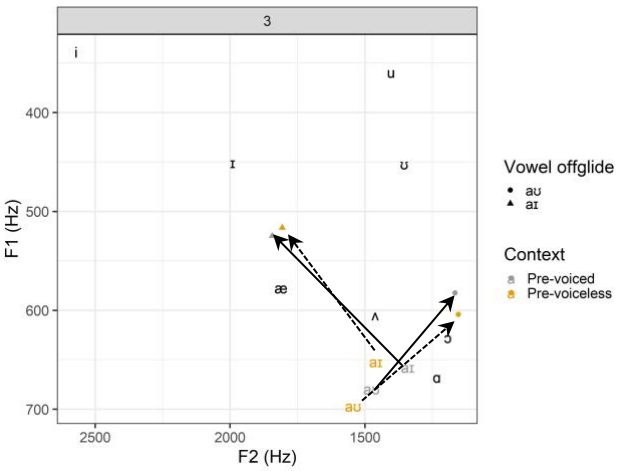


Figure 4: F1/F2 estimates of diphthong onsets and offsets in productions of a male speaker from Louisiana showing relative monophthongization (prevoiced contexts: solid lines; prevoiceless contexts: dashed lines).

It is difficult to make a robust inference for the /aʊ/ offset, however, because of the rampant variability in its acoustic values within individual speakers as well as across regions. Based on group averages (Figure 2) and individual absolute values (Table 2) and under the assumption of phonetics-phonology continuum, /ao/ (or /aɔ/ for varieties without /ɑ/-/ɔ/ merger) could have been proposed as a likely underlying form, but a close competition with /aʊ/, /au/, and in some dialects /aʌ/ casts a doubt that the assumption of two monophthong-like targets [1] would serve us well here. Furthermore, the lack of the effect of voicing in the following segment on /aʊ/'s F2 offset estimates, which is observed elsewhere in our data, strengthens the inference that the offset of this diphthong may not be a phonological target in the same sense as the offset of /aɪ/.

It is possible that these two American English diphthongs have different types of phonological representations, such as "onset plus offset" for /aɪ/ versus "onset plus direction" for /aʊ/ [3]. Such a proposal has, for example, been put forth for Ningbo Chinese, where rising diphthongs were argued to have two static targets while falling diphthongs have just one dynamic target [19]. Perhaps, at least in some varieties of English, front-gliding diphthongs like /aɪ/ might have two phonetic targets whereas back-gliding diphthongs like /aʊ/ may have a dynamic target. This proposal does not solve the issue of the most accurate IPA-style notations for diphthongs that a) do not have their offsets and onsets closely aligning to nearby monophthongs (e.g., significant differences in their positional relationship in the acoustic space) and that b) have dynamic targets that are not easily conveyed by IPA symbols.

One limitation of this study is that diphthongs' offset values are estimated as averages of the median F1/F2 values in the 80-90% portion of a diphthong, which is not an absolute endpoint of a vowel. However, F1/F2 values close to vowels' absolute endpoints are affected by co-articulation and phrasal position [11]. We assumed that the 80-90% portion of a diphthong is representative enough of the terminal acoustic target. After all, in practice, the 20-50-80% timepoints are often used for estimation of the vowel trajectory [20-21]. To check our conclusions, diphthongs in open syllables can be examined in future research, or other statistical methods can be used for the projection of the endpoints of diphthong trajectories (e.g., by estimating the amount and speed of spectral change in the last 10% of vowel duration).

Another limitation is that we took the acoustic-phonetic data to be representative of phonological targets. However, we cannot dismiss a possibility that phonological targets may be obscured by processes such as an speakers' articulatory undershot as in the hypoarticulation theory [22]. Other types of data (e.g., duration and duration ratios, kinematic data, response times in categorization tasks, rhyming trends in poetry) may provide evidence for mental representations that speakers' hold for diphthongs.

5. REFERENCES

[1] Thomas, E. R. 2020. Sociophonetic trends in studies of Southern U.S. English. *J. Acoust. Soc. Am*. 147, 529-540.

[2] Lehiste, I., & Peterson, G. E. 1961. Transitions, glides, and diphthongs. *J. Acoust. Soc. Am.* 33, 268–277.

[3] Gottfried, M., Miller, J. D., Meyer, D. J. 1993. Three approaches to the classification of American English diphthongs. *J. of Phonetics* 21, 205-229.

[4] Raffelsiefen, R., Geumann, A. 2016. AI vs. AU in American English compared to German. In: Draxler, C., Kleber, F. (eds), *Tagungsband: 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*. Ludwig-Maximilians-Universität München, 155–157.

[5] Holbrook, A., Fairbanks, G. 1962. Diphthong formants and their movements. *J. Speech Hear. Res*. 5, 38–58.

[6] Thomas, E. R. 2001. *An acoustic analysis of vowel variation in New World English*. NC Duke UP.

[7] Lee, S., Potamianos, A., Narayanan, S. 2014. Developmental acoustic study of American English diphthongs. *J. Acoust. Soc. Am.* 136, 1880–1894.

[8] Shoup, J.E. 1980. Phonological aspects of speech recognition. In: Lea, W. (ed), *Trends in Speech Recognition*. Prentice Hall, 125-138.

[9] Burgdorf, D. C., Tilsen, S. 2021. Temporal differences between high vowels and glides are more robust than spatial differences. *J. of Phonetics*, 88, 1-47.

[10] Moreton, E. 2004. Realization of the English postvocalic [voice] contrast in F1and F2. *J. of Phonetics* 32, 1–33.

[11] Pycha, A., Dahan, D. 2016. Differences in coda voicing trigger changes in gestural timing: A test case from the American English diphthong /aɪ/. *J. of Phonetics* 56, 15-37.

[12] Elvin, J., Williams, D., Escudero, P. 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *J. Acoust. Soc. Am*. 140, 576–581.

[13] Chambers, J. 1973. Canadian Raising. *Canadian Journal of Linguistics* 18, 113-135.

[14] Davis, S., Berkson, K., Eds. 2021. *American Raising.* Publication of the American Dialect Society.

[15] Moreton, E., Thomas, E. 2007. Origins of Canadian Raising in voiceless coda effects: A case study in phonologization. In: Cole, J., Hualde, J. (eds.), *Papers in Laboratory Phonology 9*. Cambridge UP, 37-64.

[16] Labov, W., Ash, S., Boberg, C. (2006). *Atlas of North American English: Phonetics, phonology, and sound change*. New York: Mouton de Gruyter.

[17] McAuliffe, M., Socolof, M., Stengel-Eskin, E., Mihuc, S., Wagner, M., Sonderegger, M. 2017-2022. *Montreal Forced Aligner* [computer program].

[18] Boersma, P., Weenink, D. 1992-2022. *Praat: Doing phonetics by computer* [computer program].

[19] Hu, F. 2013. Falling diphthongs have a dynamic target while rising diphthongs have two targets: Acoustics and articulation of the diphthong production in Ningbo Chinese. *J. Acoust. Soc. Am*. 134, 4199.

[20] Fox, R. A., Jacewicz, E. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *J. Acoust. Soc. Am*. 126, 2603–2618.

[21] Farrington, C., Kendall, T., Fridland, V. 2018. Vowel dynamics in the Southern Vowel Shift. *Am. Speech*, 93, 186–222.

[22] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W.J, Marchal, A. (Eds.), *Speech production and speech modelling*. Kluwer, 1–35.

[1] We concur with an anonymous reviewer that GAMMs may be a useful analytical tool for exploring alternative questions related to this topic. For our specific research questions, we prioritized capturing multiple formant dimensions at once.