

Vowel Dispersion in English Diphthongs: Evidence from Adult Production

Stacy Petersen
Georgetown University

1 Introduction

Although often omitted in work on vowel systems, acquisition, production and perception, diphthongs are major components of vowel inventories cross-linguistically. The decision to ignore or omit diphthongs is often due to their complex nature; a diphthong has the dual status of being a single segment while also being composed of two targets connected by a trajectory. This dual status can be problematic for unified phonological theories concerning vowels.

In this paper, I address the problem of including diphthong vowels into a Dispersion Theory (Flemming 2004) framework. First, I review the main aspects of Dispersion Theory in Flemming (2004), which gives an analysis of vowel inventories using a perception-based account of contrast, but noticeably omits diphthongs, which—while different from monophthongs—are highly productive, contrastive members of vowel inventories. Next, in order to correctly represent and incorporate diphthongs, I discuss acoustic properties of diphthongs and their presence in vowel inventories cross-linguistically. Diphthongs are compared to the monophthong inventory using production data to assess their relative positions in the vowel space. The English vowel production data should reflect the language-specific constraint ranking of *EFFORT with the maximum contrast and minimum distance constraints as predicted in Flemming's theory.

To derive diphthongs, Flemming (2004)'s constraints as well as additional constraints from Minkova & Stockwell (2003) are used to account for the distance between the two offset targets. An additional constraint is proposed to account for the strong preference in the English production data to centralize the onset targets. Derivations for individual diphthong productions compared to possible surrounding candidates are provided in the analysis.

2 Literature Review

2.1 Dispersion Framework In Flemming (2004), we see a shift from a focus on articulatory priorities in phonology to a perception-based account of contrast. To account for the goal of minimizing confusable contrasts directly in the phonology, Flemming (2004) proposes Optimality-Theoretic constraints which favor less confusable contrasts over more confusable contrasts. Due to the perceptual-based nature of this approach, the property of markedness must shift from being a property of individual sounds in an articulatory-based approach to a property of contrasts. In this analysis, the markedness of a sound is not determined inherently by its properties alone. Instead, a sound may be marked depending on the contrasts it enters into, as predicted by constraints on the distinctiveness of contrasts (235).

Flemming (2004) focuses on three functional goals that are fundamental to the selection of phonological contrasts:

- i. Maximize the distinctiveness of contrasts.
- ii. Minimize articulatory effort.
- iii. Maximize the number of contrasts.

These goals are inherently conflicting; the more distinctive a sound becomes, the more effort it takes to produce. The more contrasts there are, the less distinctive each can be. The combination and competition between these goals leads to an inventory that balances the effort, number of contrasts, and distinctiveness.

In order to formalize constraints on contrasts, a multidimensional similarity space is needed to map out the distance between the stimuli. This multidimensional map can be simplified to distinctness in two dimensions: F1 and F2. An example of a sound matrix with multiple dimensions (F1 and F2) is given in Figure 1 (Flemming 2004:238-239).

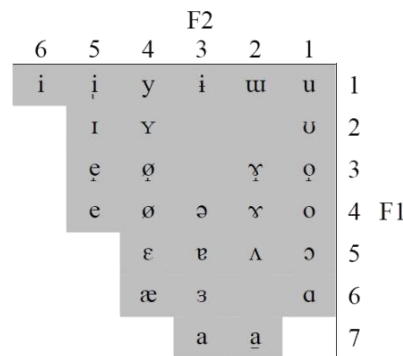


Figure 1: Multidimensional similarity space of F1 and F2

Monophthongs in the F1 and F2 dimensions in Figure 1 are given values based on their coordinates; sound distinctiveness can be calculated by the differences of pairs of vowels along these dimensions. Minimum distance constraints are in the format *Dimension : distance*. For example, MINDIST = F1:1 indicates that contrasting sounds must differ by at least 1 unit on the F1 dimension. The same method applies for F2. MINDIST constraints are inherently ranked along each dimension, e.g. MINDIST = F1:1 » MINDIST = F1:2 » ... » MINDIST = F1:4. The MINDIST constraints promote distinct contrasts and maximum dispersion in the available auditory space.

In addition to the minimum distance constraints, which promote the goal of maximizing distinctiveness in contrasts, there must also be a constraint that promotes the goal of maximizing the number of contrasts. Flemming (2004) proposes MAXIMIZE CONTRASTS, which is a positive constraint wherein a candidate is rewarded for each contrast in the inventory.

The ranking of the MINDIST constraints and MAXIMIZE CONTRASTS results in language-specific vowel inventories. Ranking MAXIMIZE CONTRASTS below MINDIST = F1:3, for example, will result in an inventory with maximum contrasts that preserves no less than 3 units of contrast between the members in the inventory. An example of this ranking is given in the Tableaux below.

	MINDIST = F1:2	MINDIST = F1:3	MAXIMIZE CONTRASTS	MINDIST = F1:4	MINDIST = F1:5
i-a			✓✓!		
☞ i-e-a			✓✓✓	**	**
i-ε-ε-a		*!*	✓✓✓✓	***	*****

Tableau 1: Example derivation of a three-vowel inventory from Flemming (2004)

There are limits on all rankings of MAXIMIZE CONTRASTS with the minimum distance constraints: not every possible ranking will result in an actual language, with none of the extreme (e.g. very high contrasts preferred) possibilities attested. No account of the effort minimization goal is given in Flemming (2004)'s analysis, but a few specific constraints minimizing effort in certain contexts are used as necessary.

MINDIST constraints can account for contrast neutralization if effort minimization is also taken into account. Neutralization in contexts where acoustic cues are weak is a pervasive characteristic in phonology (Steriade 1997, Kirchner 2004). Flemming (2004) includes the general constraint *EFFORT: if the contrast cannot satisfy a higher-ranking minimum distance constraint without violating *EFFORT, the contrast will be neutralized in a given context (see 1 below). It follows that neutralization occurs in contexts with weak cues because it will take too much effort on the part of the speaker to reach the necessary dispersion level to prevent confusability.

- (1) MINDIST = *d*, *EFFORT » MAXIMIZE CONTRASTS

As Flemming (2004:243) states, the possibility of realizing a contrast that satisfies minimum distance constraints without violating *EFFORT is highly dependent on context, as is articulatory effort. The properties of the *EFFORT constraint are not straightforward; it is not clear if this is a categorical, gradient, or binary constraint. Flemming is hesitant to completely formalize the effort minimization constraint, as it is very dependent on articulatory and contextual factors which are beyond the scope of his paper.

In sum, Flemming (2004)'s analysis adapts the ranking and competition framework of OT to the goals and hypotheses of Dispersion Theory. The result of interaction of these constraints is that the contrasts are evenly distributed and as far apart as possible in the vowel space. However, this analysis can only account for monophthong vowel inventories; ideally, vowel inventories with diphthongs should be included in this framework to create a unified theory of vowel dispersion.

2.2 The Nature of Diphthongs This section provides a brief description of the acoustic properties of diphthongs and their role in vowel inventories. Onset and offset targets are points of measurement at either end of a diphthong. These points are not necessarily taken during the steady state, as previous research (Gay 1968, Borzone de Manrique 1979) has shown that steady states are inconsistent across speech rates, different languages, and across diphthongs themselves. More precisely, they are point measurements taken directly after the perseveratory consonant transition into the diphthong and directly before the anticipatory consonant transition out of the diphthong. Production studies (Holbrook & Fairbanks 1962, Lehiste 1964, Lee et al. 2014) have shown that the onset and offset target formant values are different from formant value measurements of comparable monophthongs and/or semi-vowels. For this reason, production data was gathered for the present analysis.

A trajectory is the connecting movement between the two targets of a diphthong. The trajectory is often measured by the slope of F2 which is the rate of change in Hz (cycles per second) over a given duration of time. The trajectory is commonly measured by hand (Gay 1968, Borzone de Manrique 1976, Jha 1985, Aguilar 1999), but may be measured using computational algorithm (Dolan & Mimori 1986). A comparison of the perceptual importance of onset + offset targets vs. trajectory to diphthong identification is reviewed in Section 2.3.

Later literature moved beyond the specific phonetic characteristics of diphthongs to address them in a broader context in vowel systems and cross-linguistically. In his chapter on phonetic universals in vowel systems, Lindblom (1986) finds that, according to typological data, diphthongs that have a greater distance (trajectory) between the two targets are preferred. He provides the following hierarchy: [aj, aw] » [ej, ow] » [uj, iw]. Miret (1998) examined cross-linguistic data for frequency and combinatorial patterns. Similar to Lindblom (1986), Miret (1998) concludes that the most important factor for possible diphthongs is the difference in sonority of the components: diphthongs tend not to have two sounds of equal sonority because if they can both be nuclei in a syllable, they may be mistaken for hiatus. Frequency data supports this, as combinations of low vowels in diphthongs are not well attested. The typology also suggests that changes in both backness and height, as opposed to height or backness changes alone, between the two diphthong targets leads to maximal differentiation and diphthongs with both height and backness differences are typologically more frequent. Interestingly, Miret (1998)'s conclusions that distance between the members of the diphthong is necessary for perceptual purposes suggests diphthongs can be well adapted into a system of perceptual contrasts such as Flemming (2004)'s Dispersion Theory.

One typological study which does not reflect these findings is Maddieson (1984)'s *Patterns of Sounds*. Maddieson found that languages prefer diphthongs that begin or end with a high vowel; this supports Miret's findings that combinations of low vowels are dispreferred. However, Maddieson's findings contradict Miret and Lindblom in that maximizing the distinctiveness between targets does not explain diphthong typology, as diphthongs with short trajectories are among the most common types of diphthongs.

Although the present study focuses primarily on English, it is clear that the literature on cross-linguistic preferences in diphthongs is not agreed upon. This may be due to various factors, including the criteria these typological studies used to differentiate between diphthongs, hiatus, offgliding of monophthongs, etc. In sum, it appears that the difference between the two targets is important, but whether it is maximum distance or merely sonority which plays a bigger role perceptually remains to be seen.

2.3 Diphthong Perception It is crucial to examine existing literature on the perception of diphthongs, as Flemming (2004)'s Dispersion Theory relies on perception-based OT constraints competing with articulatory effort minimization constraints to create attested vowel inventories. Therefore, to fit diphthong

perception and production into Dispersion Theory, we must know what cues listeners are relying on to identify diphthongs. The two main components listeners may be using are the onset + offset targets and/or the transition between the targets.

Gay (1970)'s study of American English diphthongs pits duration cues against target frequency cues to explore perception of diphthongs as well as their phonemic boundaries. Gay (1970) found, using synthetic speech tests, that the primary feature of [AW, AY, ɔi] is the transitional duration rather than change in onset and offset target points. These results suggest using Flemming's Dispersion matrix may not account for diphthong perceptual data from an explanatory standpoint. Morrison (2013) criticizes Gay's methodology, stating that the synthetic stimuli used in Gay (1970) confounded offset and slope or duration and slope, leading to unclear results.

The onset and offset target points have been shown to be relevant perceptual cues to both monophthong and diphthong identification (Bladon 1985, Gottfried et al. 1993, Pitermann 2000, Morrison 2013). Gottfried et al. (1993) examines the three main approaches to characterizing diphthongs by evaluating their performance in a statistical pattern recognition task. In this way, the hypothesis that defines diphthongs with the most acoustically relevant properties (onset + offset, onset + slope, or onset + direction) which produces the best parameters for Bayesian classification of diphthongs, even under varying conditions of tempo and stress, is identified. American English diphthong data was recorded from 4 speakers at two rates and with two different stress patterns. These hypotheses were evaluated by means of a Bayesian classifier, which uses the statistical properties of classes of diphthongs to classify tokens. The results show that while each of the three hypotheses yielded very accurate results (>90%), the highest percentages that were obtained were for the onset + offset hypothesis, which averaged at 96% correct classification. The findings of this study suggest an analysis in Flemming (2004)'s framework is possible, as onset and offset points (rather than transitional data or duration data) are relevant for perceptual purposes.

2.4 Diphthongs in OT The literature in this area is very limited. Included here are two studies which address diphthongs in a broader phonological context and give insight to what an optimal diphthong vowel inventory should look like. These papers also provide potential OT constraints to use in the present analysis.

The analysis of diphthongs in OT is not the primary objective in Bermúdez-Otero (2003); however, the author does use evidence of diphthong raising and flapping in Canadian English counterbleeding as support for a Stratal OT model. Crucially, this analysis proposes the constraint CLEARDIPH, a context-free markedness constraint that favors diphthongs with maximum auditory distance between the onset/offset targets. Bermúdez-Otero (2003) also proposes a context-sensitive markedness constraint CLIPDIPH, which demands a minimization of the distance between the two targets. These two markedness constraints are in direct opposition, and it is not clear why both should be present in the constraint inventory; additionally, neither are grounded nor motivated outside the data set given. I follow the context-free versions of these constraints used in Minkova & Stockwell (2003) and Flemming (2004), *EFFORT.

The main purpose of Minkova & Stockwell (2003) is to derive English vowel shifts in bimoraic nuclei (nucleus-glide dissimilation, nucleus-glide assimilation, chain shift, and merger). They argue that this phonological restructuring is due to competing phonetic and phonological goals to create 'optimal' diphthongs. For Minkova & Stockwell, 'optimal' diphthongs have the maximum distance ($\Delta F1$ and $\Delta F2$) between the two targets. Minkova & Stockwell (2003) provide relevant constraints for diphthong analysis, HEARCLEAR and *EFFORT, both which function similarly to CLEARDIPH and CLIPDIPH, with the exception that they are both context-free. HEARCLEAR is rooted in the goal of perceptual distance, based on the F1 (height) and F2 (backness) parameters, between the onset and offset targets:

- (2) HEARCLEAR: Maximize the auditory distance between the nuclear vowel and the following glide (measured in formant frequency)

This constraint is grounded in the findings that vowel inventories seek to maximize the distinctiveness of contrasts (Flemming 2004, Lindblom 1986, Miret 1998).¹ Neutralization in contexts where acoustic cues are weak is a pervasive characteristic in phonology (Steriade 1997).

Deriving their methodology from Flemming's earlier work (1995a), Minkova & Stockwell's HEARCLEAR is separated into its two parameters: HEARCLEAR F1 and HEARCLEAR F2; determining violations follows a similar framework Flemming's (1995a, 2004) matrix. Minkova & Stockwell (2003) also follow Flemming (1995a) in including the MINDIST constraints. Diphthong candidates are evaluated one at a time rather than the entire diphthong inventory at once. This is due to the fact that diphthongs have two targets rather than one; evaluating the inventory as a whole is left to future work.

3 Data

3.1 Data Collection Recordings were made in a quiet room from three female participants who read a four-paragraph pre-prepared script. The script was a modified version of "Arthur the Rat," a well-known script originally from Henry Sweet (1890), often used by phoneticians to capture different vowels and consonants in different environments. The text was modified by inserting words and small phrases to include more tokens of diphthong vowels (total words = ~358). The average duration of each passage was 100 ±4 ms, indicating all the reading speeds were equivalent.

The recordings made in this study were made on a Sony IC Recorder, set to a bit rate of 192 kbps and a sampling rate of 44.1k Hz. Word tokens were extracted using Praat; points for vowel formant measurements were made in corresponding textgrid files, which were run through a Praat script to extract F1, F2, and F3 formant values. Data which appeared to be outliers were double checked by hand and the formant values were re-adjusted if necessary (i.e. if the script was not accurate due to an anomaly in the spectrogram). Vowel formant plots were made using online normalization software: NORM Vowel Normalization and Plotting Suite.² All plots are normalized using the Lobanov method to reduce variation caused by physiological differences between speakers.

For each diphthong, a measurement was taken at two target points: the first at the onset of the diphthong and the second at the offset of the diphthong. The first measurement point is at the furthest left point of the diphthong without including formant transitions from the preceding consonant. The second measurement point is at the peak of the offset to the rightmost of the vowel excepting any following formant transitions due to the following consonant. The decision of where to measure these two points was largely due to descriptions of the relative importance of the endpoints of the diphthong compared to the actual trajectory in the literature.

Monophthong data was collected and analyzed in Petersen (2014) to test whether Flemming (2004)'s analysis could derive a vowel space based on production rather than from the idealized similarity space map. Petersen (2014) used the constraints in Flemming (2004) to account for the position of produced monophthongs.

3.2 Data analysis Below, Table 1 provides the specific measurements for phonemic diphthongs [AY, AW, OY] and phonetic diphthongs [EI, OW]. An overall total of 200 tokens are included.

Vowel	Speaker	# tokens	Avg. F1 onset	Avg. F2 onset	Avg. F3 onset	Avg. F1 offset	Avg. F2 offset	Avg. F3 offset
AY	TOTAL:	54	791	1660	2940	618	2258	2970
AW	TOTAL:	51	771	1824	2862	741	1537	2873
EY	TOTAL:	40	576	1922	2904	509	2548	3171
OY	TOTAL:	18	575	1192	2859	528	2276	2907
OW	TOTAL:	37	607	1713	2868	608	1285	2939

Table 1: Diphthong data of three speakers

¹ Note that Maddieson (1984)'s typological results do not support these findings. However, the majority of literature on the topic does support maximum distance constraints.

² NORM vowel normalization and plotting software is not compatible with IPA symbols. Arpabet symbols are used for this purpose.

The diphthong formant measurements are plotted in Figure 2, below. The relative distance between onset and offset targets (length), slope, and location in the vowel space of the diphthongs (both onset and offset targets) are consistent between the speakers. One minor difference appears to be variation in height, rather than backness. As the speakers are consistent, Figure 2b shows the averages of the diphthongs.

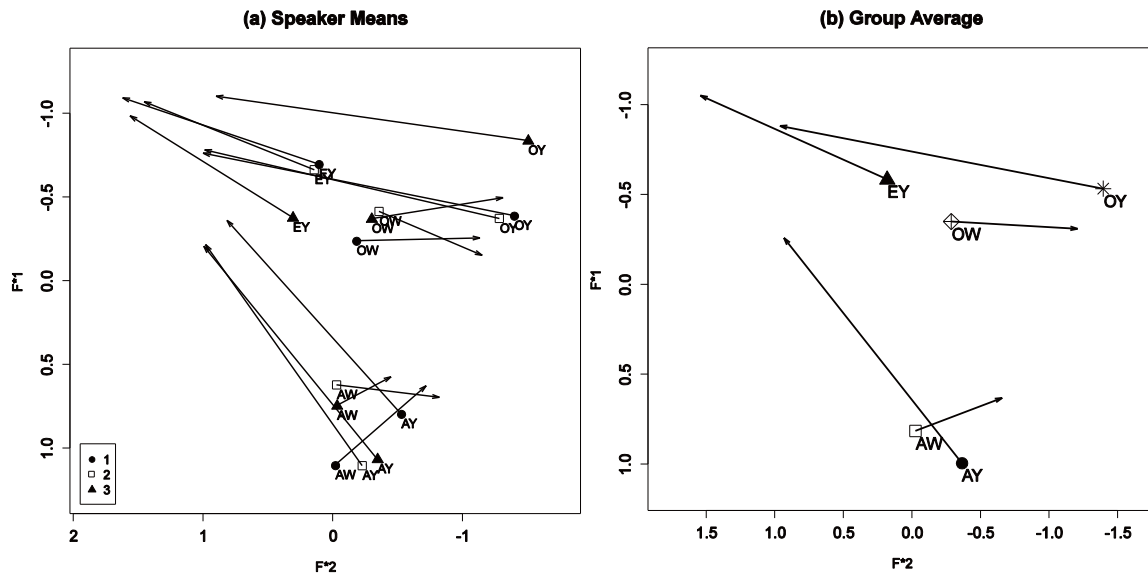


Figure 2: (a) Diphthongs by speaker means; (b) Diphthongs averaged over group

4 Analysis

In this section, the production data collected will be incorporated into an analysis of diphthongs in the Flemming (2004) Dispersion Theory vowel inventory framework. The two aspects of this study address two main omissions in Flemming's work on vowel inventories. The first aspect is that Flemming's (2004) similarity space is an idealized version of what the vowel space should look like, with very strict segmentation of the space into the corresponding vowels. In addressing vowel inventories, Flemming does not base inventories on actual production; instead, he assumes production matches the common IPA transcriptions used for a language's inventory. The usage of true production data addresses this first issue, and Petersen (2014) shows that it is possible to derive the English monophthong vowel inventory as it reflects production. The second aspect addressed is the omission of diphthongs from Flemming's vowel inventories, despite diphthongs being common productive members of vowel inventories in a large number of languages cross-linguistically.

4.1 Production data vs. Flemming (2004) similarity space Recall Flemming (2004) adopts a vowel space which displays the maximum amount of distinctions possible for vowels along the two dimensions F1 and F2; the vowel space is split up into a 6X7 grid from which coordinates are used to specify sounds in the matrix (eg. (F1=1, F2=6) for [IY]). Vowel distinctiveness can then be measured by the difference on the F1 and F2 dimensions. The vowel space in Figure 3 is from Flemming (2004:239), with the vowels of English³ circled and diphthongs indicated with arrows (a dashed circle indicates the vowel is present as an onset of a diphthong).

With the exception of the central [AH] and reduced vowel [AH0], the English monophthongs are all on the periphery of the vowel space, reflecting the perceptual goal to maximize the distinctiveness of contrasts. English also has a large number of vowels in its inventory, which reflects the goal to maximize the number of contrasts. The central and reduced vowels are likely the result of the goal to minimize effort: making a central vowel requires the least effort; the reduced vowel is also the furthest from the edges of the vowel space.

³ Reflecting a merged set of /AA/ and /AO/ as in my participants' dialects.

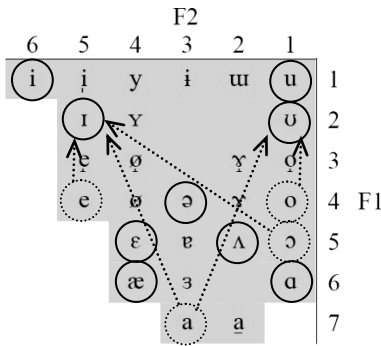


Figure 3: English vowels in Flemming (2004)'s vowel space

The diphthongs are more complicated, although they appear to prefer movement along the edges of the vowel space, with the exception of [OY] which moves through the middle. All the diphthongs appear to reflect the preference of a lower vowel for the onset target and a high (lax) vowel for the offset.

However, these circled vowels may not reflect true pronunciation by native English speakers. In Figure 4, the normalized averages of the English diphthongs and English monophthongs from Petersen (2014) are superimposed on the Flemming (2004) similarity space. To create this plot, I overlaid the monophthong and diphthong averages and adjusted them to scale while maintaining the aspect ratio of both the F1/F2 plot and Flemming's matrix. I principally matched the locations of [IY] and [AA], which are at opposite ends of the space, to determine the correct positioning of backness and height. Recall that the symbols used for the production data are not IPA; for ease of the reader, a code legend is provided to the right of the graph. A separate plot of just the monophthong overlay⁴ is provided in the Appendix. Circles around the monophthong vowels are included.

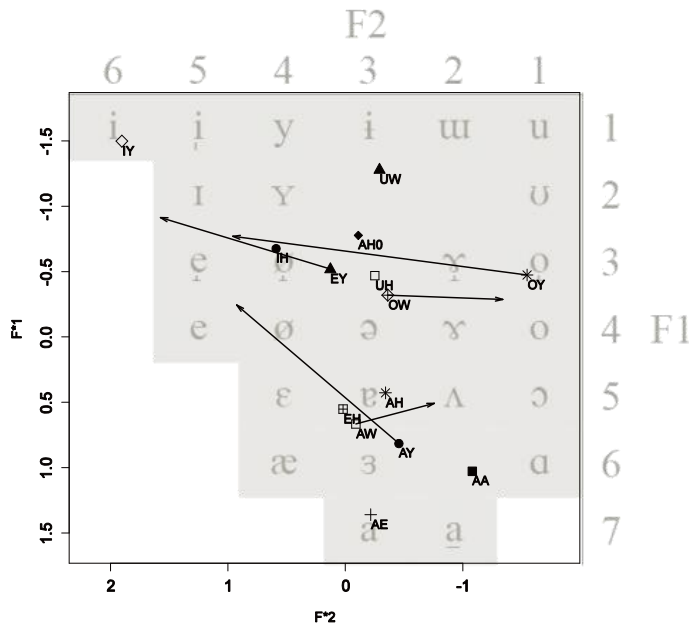


Figure 4: English monophthong and diphthong data superimposed on Flemming (2004) vowel space

Actual monophthong and diphthong production does not align with how they are presented by Flemming (2004). For diphthongs, production does not align with the monophthongs they are transcribed with. This corresponds with observations made as early as 1961, with Lehiste & Petersen's statement, "Neither of the

⁴ A plot with only the diphthongs would be impossible to align, as there is no reference to where the diphthongs correspond to without the monophthongs as reference points. See Petersen (2014) for the monophthong analysis.

elements comprising the diphthong is ordinarily phonetically identifiable with any stressed English monophthong; for example, in /AY/ the first element is neither /AA/ nor /AE/, and the second element is neither /IY/ nor /IH/ (276).

From Figure 4, it is clear that the actual production of diphthongs do not have the length as predicted in Figure 3, which shows what diphthongs would look like if their onsets and offsets correspond to their IPA labels and monophthong counterparts. The distance between the two targets can be calculated by their Euclidean distance, assuming the two points have Cartesian coordinates (x_1, y_1) and (x_2, y_2) , using the formula in (3):

(3) Euclidean distance between two points

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

For example, the modeled diphthong [AY] in Figure 3 has the coordinates, in the form of (F2, F1), of (3, 7), (5, 2). The Euclidean distance for [AY] is 5.4. The distance measurements for the diphthongs in Figure 3 and the actual distance measurements, rounded to nearest tenth, in Figure 4 are given in Table 2.

Diphthong	Fig. 13	Fig. 14
AY	5.4	4.2
AW	5.4	1
OY	5	4.1
EY	2	2.2
OW	2	2

Table 2: Euclidean distance of diphthongs

The Euclidean distance is shorter in all the 'true' phonemic diphthong pronunciations than where they would be placed by their IPA labels in Flemming (2004)'s similarity space. The two phonetic diphthongs, [EY] and [OW], however, have equal or longer distances than in Figure 3. The difference, however, is that the production data shows movement along the F2 dimension rather than the F1 for these two diphthongs. [AW] is the most reduced of all, with movement even less than the phonetic diphthongs, at only 1 unit of distance. The diphthongs [AY] and [OY] have the longest distances and move in similar trajectories to Figure 3.

Finally, all the studies cited in Morrison (2013) agree that the onset target's formant values are perceptually important cues to the vowel identification, yet are in disagreement what additional cues are relevant. He found that onset + offset formant measurements provided the best perceptual model of vowel movements. The production data shows, however, that onset formants are the least accurate targets for all the diphthongs except [OY]. The onset targets that are [-low] ([EY, OW]) cluster around the reduced vowel [AH0]. Onset targets that are [+low] ([AA]) cluster around the mid central vowel [AH]. This evidence contradicts the previous literature in that onsets do not appear to be maximally distinct nor do their formant values reflect accurately the onset vowel quality. Because the onsets are reduced, listeners may not be using onset cues for vowel identification.

The offset targets are generally more accurate. Two out of three diphthongs that end in [IH] manage to reach F2=5, F1=2 in Flemming's matrix and also surpass the monophthong average production of [IH]; this may suggest that these onsets should perhaps be transcribed at [IY] rather than [IH]. [AY]'s offset target is only off by a Euclidean distance of 1.4 for the matrix and 1 from the monophthong production of [IH]. The offset of [AW] is furthest from Flemming (2004)'s matrix position of [UH] and the monophthong production of [UH] (by 3.2 and 2.2 Euclidean distances, respectively). Although [OW]'s onset target is centralized, its offset target is only 1 unit away from Flemming's matrix [UH].

In sum, there is much more reduction of monophthong and diphthong vowels in the production data than a strict matrix (like Flemming (2004)'s similarity space) predicts. The onset targets are not maximally distinct between themselves; instead, they tend to cluster around the higher reduced vowel [AH0] and lower central vowel [AH]. Between onset and offset targets, the phonetic diphthongs show the closest resemblance to the Flemming matrix. With the exception of [AW], the phonemic diphthongs do have a large Euclidean distance between the onset and offset targets, although not as much as in the matrix. The exception, [AW] appears to be completely reduced.

4.2 OT Analysis In this section I derive the individual diphthongs as they are pronounced compared to losing forms (ie. the vowels within diphthongs rather than diphthongs compared to each other in the vowel space). The symbols I use in this section correspond to the IPA in Flemming's similarity space. Where there is no symbol provided, I note the symbol to be used for that position in the matrix. The winning candidates are diphthongs in the positions as indicated by the production data.

For this analysis, it is necessary to include additional constraints. The first additional constraint is similar to the MINDIST constraints with the exception that it is formatted specifically for diphthongs. Originally from Minkova & Stockwell (2003), HEARCLEAR is a constraint based on the maximization of distance between the two targets. As with the MINDIST constraints, HEARCLEAR operates on the two parameters F1 (for height) and F2 (for backness). Working against HEARCLEAR is the preference by speakers to conserve effort. In production, this occurs in two ways: decrease the distance of the trajectory, and centralize the onset target.

Decreasing the distance of the trajectory can be done with an *EFFORT constraint. Because Flemming (2004) does not fully develop *EFFORT, the version adopted here is from an unpublished work by Grosz (2006):

- (4) *EFFORTNUCLEUS = x: Assign * if the sum of the change on the F1 dimension and the change on the F2 dimension equals or surpasses x within a nucleus

This constraint essentially works by limiting the length of the diphthong without using the Euclidean distance.⁵ Similar to the MINDIST constraints, the *EFFORTNUC constraints are inherently ranked as follows:

...» *EFFORTNUCLEUS = 7 » *EFFORTNUCLEUS = 6 » *EFFORTNUCLEUS = 5 »...» *EFFORTNUCLEUS = 1

Next, I propose a constraint which appears to be a type of context-specific *EFFORT constraint as mentioned in Flemming (2004). From the monophthong and diphthong production data, it is clear that--at least in running speech--vowels tend to become centralized. For monophthongs, these are the non-corner vowels (see §5.1). In diphthongs, there is a very strong preference to centralize or reduce the onset target, which cannot be accounted for by the HEARCLEAR and *EFFORTNUC constraints alone.

- (5) REDUCEONSET =x: Assign * if the sum of the change on the F1 dimension and the change on the F2 dimension away from (F2=3, F1=4) of the onset target surpasses x
REDUCEONSET =1 » REDUCEONSET =2 » REDUCEONSET =3 » REDUCEONSET =4 » ...

The REDUCEONSET constraints only work on onsets: it forces the onset toward the center of the matrix on both the F1 and F2 dimensions. The inherent ranking of this constraint is in (5). If all candidates violate the higher ranking constraint, it is omitted for space considerations.

The following diphthong tableaux will demonstrate the interaction between these constraints to derive the actual production diphthong out of 9 possible candidates. Candidates are combinations of the closest surrounding onset and offset targets. The effect of the *EFFORTNUCLEUS constraint ranking above the HEARCLEAR constraints (which are formatted and work identically to the MINDIST constraints, except that they apply within a vowel instead of between two separate vowels) shows that the goal of minimizing effort often dominates the goal of creating the most distinct contrast possible. Irrelevant constraints are left out of the tableaux for space concerns. Omitted constraints follow inherent ranking as detailed above.

The production of [AY] has an onset target of (F2=2, F1=6) and offset target at (F2=5, F1=3). The following tableau sets the production of [AY] (here, marked as A ϵ , as Flemming 2004 does not specify an IPA symbol for the onset position) against other possible similar diphthongs that correspond to the matrix coordinates and labels in Flemming (2004)'s similarity matrix.

Tableau 2: production of [AY] ([A ϵ] in Flemming 2004)

⁵ It is possible to formulate this constraint using the Euclidean distance, except the result would have to be so precise as to include differences between fractions of numbers; the constraint here uses the pure sum rather than the root of the squared sums, which allows the distance to remain a whole number.

	REDUCE ONSET =3	*EFFORTNUC =7	HEARCLEAR =F2:3	HEARCLEAR =F2:4	HEARCLEAR =F2:5	HEARCLEAR =F1:4	HEARCLEAR =F1:5
a. Ae				*	*	*	
b. $\text{a}\underset{\cdot}{\text{e}}$			*!	*	*		*
c. $\text{a}\underset{\cdot}{\text{e}}$	*!	*		*	*		*
d. $\text{a}\underset{\cdot}{\text{e}}$	*!	*			*	*	
e. aI			*!				
f. aI	*!						
g. aI	*!						
h. ai		*!					

A higher *EFFORTNUCLEUS limit indicates the diphthong surface with a greater maximum contrast. A higher REDUCEONSET limit allows the onset to be further away from the center of the vowel space. In Tableau 2, HEARCLEAR=F2:n plays the crucial role in ruling out shorter trajectories.

The remaining diphthongs are derived in a similar manner. The constraint ranking of each is the same except for the REDUCEONSET and *EFFORTNUCLEUS limits. In Tableau 3, the symbol E is used for the coordinate (F2=3, F1=3), while U is used in (4) for the same coordinate.

Tableau 3: production of [EY] ([Eɪ in Flemming 2004)

	REDUCE ONSET =1	*EFFORTNUC =4	HEARCLEAR =F2:1	HEARCLEAR =F2:2	HEARCLEAR =F2:3	HEARCLEAR =F2:4	HEARCLEAR =F1:2
a. Ei					*	*	*
b. eI	*!		*	*	*	*	*
c. eI		*!		*	*	*	*
d. Ei		*!				*	*
e. ei	*!	*		*	*	*	*
f. ei	*!	*			*	*	*
g. EY				*!	*	*	*
h. Ej		*!			*	*	*

Tableau 3 has a much smaller REDUCEONSET limit. This causes the winning onset to be close to [AH0].

In the following tableau, $\underset{\cdot}{\text{q}}$ is used for the empty coordinate space (F2=3, F1=2). The symbol U coordinates are used for the empty space (F2=3, F1=3). Note that this space was designated with E in Tableau 3. This coordinate is also the same position of the monophthong production of the reduced vowel schwa.

Tableau 4: production of [OW] ([U $\underset{\cdot}{\text{q}}$ in Flemming 2004)

	REDUCE ONSET =1	*EFFORTNUC =3	HEARCLEAR =F2:1	HEARCLEAR =F2:2	HEARCLEAR =F2:3	HEARCLEAR =F1:1	HEARCLEAR =F1:2
a. $\text{U}\underset{\cdot}{\text{q}}$					*	*	*
b. $\text{y}\underset{\cdot}{\text{q}}$	*!			*	*	*	*
c. $\text{ə}\underset{\cdot}{\text{q}}$	*!	*			*		*
d. $\text{o}\underset{\cdot}{\text{q}}$	*!		*	*	*		*
e. oo	*!		*	*	*		
f. yU		*!		*	*		
g. $\text{y}\underset{\cdot}{\text{q}}$				*!	*		*

h. əɔ		*!			*		*
-------	--	----	--	--	---	--	---

The endpoint [ɔ] is only one unit on the F1 dimension away from [UH], but the strict ranking of REDUCEONSET causes the onset to surface closer to [AH0]. Again, HEARCLEAR=F2:n allows the longer (4a) to win over (4g). Tableaux 3 and 4 show that the constraints and constraint rankings in this analysis can account for phonetic diphthongs as well as phonemic ones.

Tableau 5: production of [OY] ([ɔi] in Flemming 2004)

	REDUCE ONSET=3	*EFFORTNUC =6	HEARCLEAR =F2:4	HEARCLEAR =F2:5	HEARCLEAR =F2:6	HEARCLEAR =F1:1	HEARCLEAR =F1:2
a. \emptyset ɔi				*	*		*
b. ɔi	*!			*	*		
c. ɔi		*!		*	*		
d. ɔi		*!		*	*		
e. ɔi		*!			*		
f. ɔɛ				*	*	*!	*
g. ɔɔ			*!	*	*		
h. ɔi		*!			*		

Tableau 4 is similar to Tableau 2 in terms of its onset and effort requirements. This is the first tableau, however, to rely on HearClear=F1:n constraints to choose the winner. Candidate (5f) does not show movement along the height parameter, causing it to lose to (5a).

Tableau 6: production of [AW] ([ɛʌ] in Flemming 2004)

	REDUCE ONSET=1	*EFFORTNUC =2	HEARCLEAR =F2:1	HEARCLEAR =F2:2	HEARCLEAR =F2:3	HEARCLEAR =F1:1	HEARCLEAR =F1:2
a. \emptyset ɛʌ				*	*	*	*
b. ɛɔ		*!			*	*	*
c. ʌ	*!	*		*	*		*
d. ɔɔ	*!	*			*		
e. ɛɔ		*!			*		*
f. ɛʌ	*!	*			*	*	*
g. ɛɻ		*!		*	*		
h. ʌ			*!	*	*		*

The production of [AW] is the least similar to its counterparts in Flemming (2004) due to its extreme reduction. Even so, the constraints and ranking below provide the correct derivation. The constraints reflect the reduction: REDUCEONSET and *EFFORTNUCLEUS have low tolerance for deviation from the central vowel onset and a short length. This is needed, as [ɛʌ] is the shortest diphthong in the inventory.

5 Conclusions

This study sought to address two main omissions in previous work on vowel inventories in Dispersion Theory. The first was to evaluate the ability of Dispersion Theory to account for actual production data. Data was collected from three native English speakers, from which recordings monophthong (in Petersen 2014) and diphthong data were extracted. Previous literature guided choices in diphthongs to be evaluated, points of measurement, and cues that are most important to diphthong perception. Data was then plotted and against Flemming (2004)'s similarity space for comparison.

In the analysis, additional constraints were introduced to account for the diphthong production data, including the inherently ranked *EFFORTNUCLEUS= x (Grosz 2006) and the constraint proposed here, *REDUCEONSET. These two additional constraints were based on articulatory goals to minimize effort: on one hand, reduced effort led to shorter diphthongs; on the other, reduced effort caused the onset target to be reduced to a central vowel position. Each derivation evaluated one diphthong amongst a set of possible candidates.

Overall, it appears that (at least in reading-speed speech) the goal for minimization of effort tends to take precedence over the goal to maximize distance in individual diphthongs. This conclusion is not entirely consistent with previous literature, which mainly states that the two targets in a diphthong seek to maximize distance. The production data even suggests that the onset target may be the least reliable cue for diphthong identification, contrary to Morrison (2013)'s study, due to its tendency toward reduction.

Future work is needed to evaluate diphthongs in comparison to one another in the vowel space. Cross-linguistic trends suggest diphthongs tend to gravitate to the vowel space periphery (Grosz 2006). An analysis of diphthong inventories with fewer diphthongs than English (which are ideally more symmetrical) would be helpful in proposing and evaluating appropriate constraints for diphthong dispersion in a vowel space. At this next stage of implementing diphthongs in Dispersion Theory, it will become evident how the goal of maximum distinctions applies to the diphthong inventory, if at all.

References

- Aguilar, Lourdes. "Hiatus and diphthong: Acoustic cues and speech situation differences." *Speech communication* 28.1 (1999): 57-74.
- Bermúdez-Otero, Ricardo. "The acquisition of phonological opacity." *Variation within Optimality Theory: Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. 2003.
- Bladon, Anthony. "Diphthongs: A case study of dynamic auditory processing." *Speech Communication* 4.1 (1985): 145-154.
- Borzone de Manrique, Ana Maria. "Acoustic analysis of the Spanish diphthongs." *Phonetica* 36.3 (1979): 194-206.
- Dolan, William B., and Yoko Mimori. "Rate-dependent variability in English and Japanese complex vowel F2 transitions." *The Journal of the Acoustical Society of America* 80.S1 (1986): S96-S96.
- Flemming, Edward. *Auditory Representations in Phonology* (Doctoral dissertation). UCLA. 1995a.
- Flemming, Edward. "Contrast and perceptual distinctiveness." *Phonetically-based phonology* (2004): 232-276.
- Gay, Thomas. "Effect of speaking rate on diphthong formant movements." *The Journal of the Acoustical Society of America* 44.6 (1968): 1570-1573.
- Gay, Thomas. "A perceptual study of American English diphthongs." *Language and Speech* 13.2 (1970): 65-88.
- Gottfried, Michael, James D. Miller, and Donald J. Meyer. "Three approaches to the classification of American English Diphthongs." *Journal of phonetics* 21.3 (1993): 205-229.
- Grosz, Patrick G. "Diphthong Inventories." (MIT, unpublished manuscript). 2006.
- Holbrook, Anthony, and Grant Fairbanks. "Diphthong formants and their movements." *Journal of Speech, Language, and Hearing Research* 5.1 (1962): 38-58.
- Jha, Sunil Kumar. "Acoustic analysis of the Maithili diphthongs." *Journal of Phonetics* 13.1 (1985): 107-115.
- Kirchner, Robert. "Consonant lenition." *Phonetically based phonology* (2004): 313-345.
- Lee, Sungbok, Alexandros Potamianos, and Shrikanth Narayanan. "Developmental acoustic study of American English diphthongs." *The Journal of the Acoustical Society of America* 136.4 (2014): 1880-1894.
- Lehiste, I. "Acoustical characteristics of selected English consonants. Research Center in Anthropology." *Folklore and Linguistics, Indiana University, Bloomington* (1964).
- Lindblom, Björn. "Phonetic universals in vowel systems." *Experimental phonology* (1986): 13-44.
- Maddieson, Ian, and Sandra Ferrari Disner. *Patterns of sounds*. Cambridge university press, (1984).
- Minkova, Donka, and Robert Stockwell. "English Vowel Shifts and 'Optimal' Diphthongs." *Optimality Theory and language change*. Springer Netherlands, 2003. 169-190.
- Miret, Fernando Sanchez. "Some reflections on the notion of diphthong." *Papers and studies in contrastive linguistics* 34 (1998): 27-51.
- Morrison, Geoffrey Stewart. "Theories of vowel inherent spectral change." *Vowel inherent spectral change*. Springer Berlin Heidelberg, 2013. 31-47.
- Petersen, Stacy J. "Vowel Dispersion in English Monophthongs: Evidence from Adult Production" (unpublished manuscript). 2014.
- Piternann, Michel. "Effect of speaking rate and contrastive stress on formant dynamics and vowel perception." *The journal of The acoustical society of America* 107.6 (2000): 3425-3437.
- Steriade, Donca. "Phonetics in phonology: the case of laryngeal neutralization." MS, UCLA 1997.