Evidence and characterization of a glide-vowel distinction in American English

Zachary Scott Jaggers

Linguistics, New York University
New York, NY, USA

zackjaggers@nyu.edu

Abstract

This study tests whether native speakers of American English exhibit a glide-vowel distinction ([j]-[i]) in a speech elicitation experiment. When reading sentences out loud, participants' pronunciations of 4 near-minimal pairs of pre-existing lexical items (e.g., *Eston*[iə] vs. *pneumon*[jə]) exhibit significant differences when acoustically measured, confirming the presence of a [j]-[i] distinction. This distinction is also found to be productively extended to the production of 20 near-minimal pairs of nonce words (e.g., *Súmia* → [sumiə] vs. *Fímya* → [fimjə]), diversified and balanced along different phonologically relevant factors of the surrounding environment. Multiple acoustic measurements are compared to test what aspects most consistently convey the distinction: F2 (frontness), F1 (height), intensity, vocalic sequence duration, transition earliness, and transition speed. This serves the purpose of documenting the distinction's acoustic phonetic realization. It also serves in the comparison of phonological representations. Multiple types of previously proposed phonological representations are considered along with the competing predictions they generate regarding the acoustic measurements performed. Results suggest that the primary and most consistent characteristic of the distinction is earliness of transition into the following vowel, with results also suggesting that the [j] glide has a greater degree of constriction. The [j] glide is found to have a significantly *less* anterior articulation, challenging the application of a representation based on place or articulator differences that would predict [j] to be *more* anterior.

Keywords: acoustic phonetics, glides, semivowels, semiconsonants, hiatus, representation, features

1. Introduction

Pre-existing lexical items suggest that a glide-vowel distinction exists in near-minimally paired environments in American English:

(1)	[CVV]:	*Estonia*	[ɛstóniə]	*millennia*	[mɪlɛ́niə],	*duet*	[duɛ́t],
	[CGV]:	*pneumonia*	[nʊmónjə]	*Kenya*	[kɛ́njə],	*dwell*	[dwɛ́l],

However, the precise nature and representation of this distinction has not yet been established. There is also a lack of phonetic documentation, which would help in deciding between the representations that have been proposed so far. While apparent in both the [j]-[i] and [w]-[u], this study focuses on the [j]-[i] distinction. Using a speech elicitation experiment, this study tests for the [j]-[i] glide-vowel distinction in American English. It also collects phonetic data along a variety of characteristics in an effort to determine the proper representation.

This study tests if the [j]-[i] distinction can be elicited in pre-existing lexical item pairs like those above. It also tests if this distinction can be productively extended to newly encountered words and elicited via <y> vs. <i> orthography. Acoustic analysis is used to capture the most consistent characteristics of the distinction. This provides documentation of the distinction, as well as guidance for how future research may best examine it. Furthermore, three competing broad classes of phonological representations that have been previously put forth are considered regarding the current analysis. This study therefore not only tests whether such a distinction is available to American English speakers; it also compares these representations, considers acoustic predictions they generate, and applies these predictions to the data at hand. This may help speak between these competing representations, by either identifying one optimal approach or at least ruling one out.

2. Background

2.1. Competing phonological representations

Competing accounts debate whether glide-vowel distinctions are phonologically possible and attested. One argument (e.g., Steriade, 1984; Kaye and Lowenstamm, 1984; Durand, 1987; Deligiorgis, 1988) is that there is no underlying distinction between vowels and glides, positing that glides are instead always phonologically derived from underlying vowels in certain environments. Levi (2004; 2008), however, provides evidence from multiple languages in which glide surface forms are not fully predictable from their surrounding environment. This unpredictability leads Levi to conclude that languages can underlyingly distinguish between glides and vowels. This study maintains the assumption (as strongly motivated by Levi, 2004) that glide-vowel distinctions are *available* to the human phonological faculty, and it tests whether such a distinction is present and

productive in the phonological system of American English. However, a further debate remains open regarding how such distinctions should be phonologically *represented*. This study therefore considers competing representation accounts and compares them as candidates for representing the apparent distinction in American English.

One kind of representation account proposes that glide-vowel distinctions are attributable to a distinction in the segment's primary articulator: a PLACE-BASED REPRESENTATION. In an analysis of multiple languages, Levi (2004; 2008) proposes that all vowels are primarily [Dorsal], while glides' primary articulators differ. Levi suggests that /w/ is primarily [Labial], while /j/ is primarily [Coronal], in accord with Halle et al.'s (2000) Revised Articulator Theory. As one example, Levi (2008) describes Pulaar as having both derived and underlyingly phonemic glides, with the derived glides being predictable by the surrounding environment while the phonemic ones are not. To argue how they are represented, Levi analyzes how the glides participate in a previously documented (Paradis, 1992) process of consonant gradation, alternating with more fortified counterparts. She demonstrates that the fortified, consonantal versions of phonemic /j/ and /w/ have coronal and labial places of articulation ([d͡ʒ] and [b], respectively), while their derived (underlyingly vocalic) counterparts fortify to a dorsal place of articulation ([g]). This representation approach is similar, though not identical, to other representations previously put forth, such as proposals that palatal [j] is both [Coronal] and [Dorsal] (e.g., Keating, 1988; Nevins and Chitoran, 2008). In terms of how such a distinction might manifest in production, we might expect tighter constriction at these more anterior places of articulation. Regarding another language, Karuk, Levi (2008) discusses how the /w/ glide is documented (Bright, 1957) to exhibit bilabial frication. And Keating's (1988) conclusion that [j] is both [Coronal] and [Dorsal] comes from X-ray analysis demonstrating coronal constriction during [j]. Therefore, in the analysis at hand, the primary articulation of the [j] glide would be predicted by this account to be more anterior than that of its [i] vowel counterpart. In terms of constraints on the phonological distribution, such a representation might predict this distinction to exhibit homorganicity effects and be constrained by the place of articulation of surrounding sounds.

Another kind of account, henceforth referred to as a CONSTRICTION-BASED REPRESENTATION, makes use of the notion of constriction degree to distinguish glides and vowels, positing that the production of glides involves tighter constriction of the vocal tract than the production

of their vowel counterparts. Padgett (2008) proposes that in systems with a glide-vowel contrast there is no difference of articulator or frontness of articulation, but only that of the feature [±vocalic]. Padgett proposes that glides, being [–vocalic], have a distinctly tighter degree of constriction than their [+vocalic] counterparts, equal along all other featural dimensions. One of the examples Padgett provides is the previously documented (Townsend and Janda, 1996) pattern of Slavic stops mutating into palatalized, affricated counterparts. This pattern is more frequent when a stop is followed by a glide than when followed by a vowel, and more frequent when a vowel is high than when it is not. Padgett argues that this scale of likelihood is attributable to the degree of constriction of the following segment: the narrower a following segment, when coupled with the release of the stop, the more likely the release is to be perceived and reanalyzed as affrication; therefore, the glide is narrower than its high vowel counterpart. This is similar to previous proposals suggesting that glides have narrower constriction targets (e.g., Straka, 1964; Maddieson and Emmorey, 1985), with arguments referring both to phonological distribution and phonetic properties such as a lower acoustic intensity of glides. Such proposals vary, however, in precise feature specification, such as employing [±consonantal] instead of [±vocalic] (e.g., Hyman, 1985; Hayes, 1989; Rosenthall, 1994). For the distinction of interest in this study, such an account would entail that the production of a lingual [j] glide has a tighter constriction reached by a higher lingual articulation than its [i] vowel counterpart. In terms of distribution, this representation might predict the factor of sonority or openness to play a role in constraining the distinction.

Another kind of account is what will be referred to as a SYLLABIC PRE-LINKING account. Levi, in her cross-linguistic documentation and analysis (2004), maintains that two different types of glide-vowel distinctions are typologically possible. One is the place-based representation as introduced above. The other is based on Levin's (1985) notion of "pre-linking" with respect to syllabification. In this type of system, glide and vowel counterparts are identically featured with respect to both place and constriction. Instead, identical segments in the lexical representation can be anticipatorily specified in terms of how they will be syllabified. While Levi (2004) concludes in favor of a place-based representation for the distinctions observed in some langauges, she concludes that an apparent distinction in Spanish is best analyzed as a syllabic pre-linking kind of system, with some cases of vowels exceptionally specified in the lexeme's underlying form

to surface as vocalic syllable nuclei in environments where Spanish phonology more commonly dictates them to surface as their non-nuclear glide counterparts. This is in line with other previous accounts (e.g., Roca, 1997; Harris and Kaisse, 1999) suggesting that the [GV]-[VV] distinction is a result of underlyingly specified syllabification of identically featured vowel phonemes. In the case at hand, both [j] and [i] surface forms would be underlyingly /i/ and an [iV] hiatus output would be the result of the underlying /i/ having been pre-specified to surface as a syllable nucleus, therefore not being parsed into the syllable margin as might otherwise be allowed or preferred by the language's phonotactics. Regarding production, due to this difference in syllabification such a distinction might be predicted to manifest mainly via timing differences, with a glide being shorter than its vowel counterpart but not of a differently specified target or degree of constriction. Along these lines, Catford (1977:165) argues that glides are intrinsically fast and dynamic without a "noticeable duration" like their vowel counterparts (cf. Maddieson, 2008), but identical in terms of "articulatory stricture". Regarding distribution, Levi (2004) suggests that such a distinction is more idiosyncratic and marked, and that the vocalic option is underrepresented across the lexicon.

The next section (§2.2) will discuss where glide-vowel distinctions appear available or constrained in the American English phonology. However, these observations regarding the distribution do not strongly speak to which kind of representation might be best applicable to the case at hand, therefore motivating analysis of production and how it may speak between the competing accounts.

2.2. Considering the phonological distribution in American English

The [j] glide in American English can occur as a simplex word-initial onset, followed by a host of following vowels: e.g., *yolk* [jok], *year* [jiɹ], *Yale* [jel], *young* [jʌŋ]. However, we rarely encounter word-initial cases of [iV] hiatus, except a few cases in which the initial segment is stressed (e.g., *eon* [íɑn], *Ian* [íən]); the glide-vowel distinction may therefore not be available in this environment without being confounded with stress.[1] Following word-initial consonants, the distinction appears to be available: pairs like *fjord* [fjɔɹd] + *Fiona* [fiónə] and *dwell* [dwɛl] + *duet* [duét] exhibit the [j]-[i] and [w]-[u] distinctions, respectively. This study therefore limits the analysis

to environments with a preceding consonant (C_V), where the distinction between a glide and its unstressed vowel counterpart does seem to be available. Another narrowing of the scope of this analysis regards [ju]. This study focuses only on [j] when it is its own glide segment, rather than part of a diphthongal nucleus. There is a large and varied host of [CjV] items in which the following vowel is [u]: e.g., *fume* [fjum], *huge* [hjud͡ʒ], *cute* [kjut]. However, previous research has shown that /ju/ appears to pattern as a monomoraic diphthong in the English phonemic inventory (Jensen, 1993; Davis and Hammond, 1995), in which the high front vocoid behaves differently from the glide of interest here. And Smith (2003) documents that nuclear vs. onset onglides can exhibit different phonological behavior. This further supports treating [j] as distinct from the glide in the [ju] diphthong. Throughout this paper, the [j] glide of interest is that which is not a member of the [ju] diphthong unless otherwise noted.

This section discusses the apparent distribution of glide-vowel distinctions in American English, narrowing in on the C_V environment where a glide-vowel distinction does seem to be available. There are further constraints apparent regarding the distribution of such a distinction, which may speak to our consideration of competing phonological representations. However, it is important to note that this distinction does not appear to be robustly prevalent across the English lexicon—an aspect which (according to Levi, 2004) would lend support to applying the syllabic pre-linking representation—and is not well-documented. It is also variable across words and speakers. Therefore, the reported surface forms of different words considered here are most certainly not meant to be presented as categorical. But, they have been cross-checked with the reports of other linguists and the transcriptions provided by multiple online dictionary sources: e.g., Dictionary.com (dictionary.com), Merriam-Webster (merriamwebster.com), Cambridge Dictionary of American English (http://dictionary.cambridge.org/us/dictionary/english/). While some dictionary entries acknowledge potential for variation between [jV] and [iV] pronunciation, some list only one pronunciation and across them a majority vote can become apparent. A final note before proceeding is that, when considering the phonological distribution of glides and vowels in this section, the [w]-[u] paradigm is also taken into account simply to show that the distributions seem similarly constrained across the two paradigms. However, as has already been made clear, the experimental analysis conducted in this study focuses on the [j]-[i] distinction. (This is largely due to complications that would arise in trying to apply acoustic phonetic analysis alone, as pursued

in this study, to aptly describe the articulation: given the labial gesture, additional ultrasound analysis (e.g., Gick, 2002; Stone, 2005; Davidson, 2006) of lingual constriction and positioning would be ideal to provide a full description of how [w] and [u] may be distinctly produced.)

One apparent constraint is that of the place features of neighboring segments. First regarding the [w]-[u] paradigm, there seem to be no cases of [CwV] in which the following vowel or diphthong involves a high back vocoid: e.g., *[Cwaʊ], *[Cwu] (Davis and Hammond, 1995). The appearance of [w] in [CwV] sequences also seems to avoid labial preceding consonants: e.g., *[fwV] (Clements and Keyser, 1983).[2] While [w] is considered labiovelar, [CwV] sequences with dorsal preceding consonants are not banned: e.g., *quit* [kwɪt], *awkward* [ɑ́kwɚ˞d]. Regarding the [j]-[i], there are many fewer cases of [CjV] sequences. In fact, Davis and Hammond's (1995) paper on "onglides in American English" doesn't mention non-[ju] cases of [j] when discussing post-consonantal environments. However, some near-minimal pairs do show a distinction between [i] and [j]: e.g., *Estonia* [ɛstóniə] + *pneumonia* [nʊmónjə]. Many of these cases are variable, but they do show trends analogous to those apparent regarding [w]. No such cases are ever followed by a nucleus containing a high front vocoid such as *[Cjaɪ] or *[Cji]. Regarding place of articulation of the preceding consonant, the case of *fjord* [fjɔɹd] suggests that preceding labial consonants do not prohibit the glide. A more established word in English, *piano*, also allows for [j] in surface form [pjǽno] (while exhibiting inter-speaker variability with [piǽno]). In cases where we might expect [j] to appear following a dorsal consonant, such as the borrowing of placename *Kyoto* (Japanese source form [kʲoːto]), the adaptation instead appears to prefer a full [i] vowel in American English: [kióɾo]. We also do not see word-initial [CjV] sequences in which the preceding consonant is coronal. These observed homorganicity constraints could support a place-based representation, with a possible interpretation being that both restrictions against preceding dorsal and coronal consonants are due to homorganicity and that /j/ is therefore underlyingly both coronal and dorsal (Keating, 1988; Nevins and Chitoran, 2008; cf. Levi, 2008).

The sonority of the preceding segment also appears to constrain the distribution of glides. Take, for example, the loanword adaptation of French *noir* (source form [nwaʁ]). While [w] is allowed after other coronal consonants (e.g., *dwarf* [dwɔɹf]), *noir* is commonly adapted to [nuáɹ]. Following Steriade (1988), homorganicity constraints can play a role in the interactions of segments both within the syllable margin and between the margin and the nucleus (as observed

above), while sonority constraints only play a role within the syllable margin but not between the margin and the nucleus. Therefore, the adaptation of *noir* banning a *[nw] onset while allowing [tw], [dw], and [sw] onsets could be attributable to the constraint against the flatter sonority cline within that complex onset. And [j] appears to pattern in parallel, with no clear cases of a word-initial *[mj] onset. (The fact that *music* invariably maintains the form [mjúzɪk] is one argument in support of the glide in [ju] pertaining to a diphthongal nucleus, suggesting that the nasal and glide are not both within the syllable margin where the constraint against a flatter sonority cline would be applicable.)

Turning attention to word-medial position, examples suggest that both homorganicity and sonority constraints may be circumvented: *Kenya* [kɛ́njə], *pneumonia* [nʊmónjə]. This seems due to the option of licitly syllabifying the preceding consonant to the coda of the preceding syllable: [kɛ́n.jə], [nʊ.món.jə]. However, this syllabification appears to be constrained by the sonority of the preceding consonant. In the adaptation of placename *Tokyo*, in which a glide adaptation might be expected as a more faithful replication of the Japanese source form [to:kʲo:], the homorganicity constraint regarding the preceding consonant seems to reappear, resulting in a full vocalic adaptation: [tó.ki.o]. Following Gouskova (2004), this may be due to what would otherwise be too steep a rise in sonority across the syllable boundary: *[tók.jo]. Avoiding coda syllabification of the voiceless stop, the homorganicity constraint then applies, which disprefers the complex *[kj] onset in the *[tó.kjo] adaptation candidate and leads the [tó.ki.o] candidate to be the winner. The observation that sonority cline constraints may play a role in the distribution of this distinction could lend support to a constriction-based proposal that glides are underlyingly specifed as less sonorous than their vowel counterparts.

While we can further understand the constraints on this distinction by analyzing its phonological distribution, this does not present a clear choice between the competing place and constriciton/height representations proposed, as both homorganicity and sonority appear to play a role. While not speaking between these two representations, the observation of both of these effects could therefore lend support to a hybrid account like that by Nevins and Chitoran (2008): that [j] differs from [i] along both the dimensions of place/ariculator ([Cor, Dors] vs. just [Dors]) and constriction ([–vocalic] vs. [+vocalic]). However, as also acknowledged in this section's discussion, this distinction appears to be somewhat infrequent across the lexicon, as well as vari-

able. This could lend support to treating the current case as one of lexically exceptional syllabic pre-linking (Levi, 2004). The following section (§2.3) will turn to discussing how each of these representations generates different predictions in the articulation and, therefore, acoustics of the [j]-[i] distinction of interest here, motivating the acoustic analysis pursued in this study.

2.3. Acoustic characterization

The competing phonological representations considered above generate different predictions regarding the acoustics of the [j]-[i] glide-vowel distinction of interest in this study. There are acoustic aspects widely considered to correlate with lingual frontness, height, and constriction. There are also aspects related to timing that may play a crucial role in such a distinction across the different representations, though arguably the only role in the syllabic pre-linking account. Therefore, analyzing the acoustics of such a distinction may speak to which representation appears more directly borne out in production, or at least rule one out.

Recall that a PLACE-BASED REPRESENTATION suggests that /j/ is primarily [Coronal] (or at least includes a [Coronal] specification), while /i/ (like all vowels) is [Dorsal]. This predicts that a [j] production should have more fronted tongue mass than [i] if the primary articulator and/or target of constriction is anatomically more anterior. Acoustically, we can analyze F2 to examine a vocoid's frontness: more anterior vocoids have a higher F2 than more posterior ones. This representation therefore predicts that [j] should reach a higher F2 than [i].

On the other hand, a CONSTRICTION-BASED REPRESENTATION suggests that /j/ does not differ at all in place from /i/. Instead, it contends that glides have a tighter constriction than their vowel counterparts. In this case, then, [j] production should involve a tighter constriction achieved by a higher lingual articulation. Two acoustic measurements may capture this. The first is F1, which can be used to examine a vocoid's height: vocoids of a higher lingual articulation have a lower F1 than vocoids of a lower lingual articulation. This representation therefore predicts that [j] should reach a lower F1 than [i]. It also predicts that [j] should have a lower acoustic intensity due to its narrower constriction of the vocal tract.

There is acoustic documentation suggesting that we may find such acoustic aspects to characterize the distinction at hand. In studies of intervocalic glides (i.e., [VGV] environments), a

dip in intensity between the first and second vowel is observed when an intervening glide is present, as compared to [VV] hiatus (e.g., Aguilar, 1999; Davidson and Erker, 2014). However, especially given Straka's (1964) finding that there does not seem to be some consistent threshold across phonological contexts, this could differ in the environment of interest here: [C_V]. And in an acoustic phonetic analysis of glides and other approximants in English, Espy-Wilson (1992) finds [j] to correlate with a higher F2 and lower F1 than [i], therefore potentially supporting a combination of the place- and constriction-based representations (like that put forth by Nevins and Chitoran, 2008). However, that study does not directly examine this potential underlying distinction in controlled and balanced environments but the comparative acoustics of differently identified surface forms.

In a SYLLABIC PRE-LINKING account, only timing should play a role in conveying such a distinction since no difference is proposed regarding place or constriction. One way is in the duration of the entire vocalic sequence. A [jV] sequence may be shorter overall than a [iV] sequence due to the prior being one syllable instead of two. Crystal and House (1990) observe that the number of syllables within a stress group can influence its duration: more syllables means a longer duration. They also find that syllables with fewer segments tend to be shorter. And while a [iV] sequence has more syllables (two as opposed to one), those syllables each have fewer segments. However, the effect size observed by Crystal and House is greater for higher prosodic units. Therefore, based on the number of syllables (vs. number of phones per syllable), the prediction stands that a [jV] sequence's duration will be shorter than that of a [iV] hiatus sequence.

Another timing factor that could play a role is the speed of transition. We might expect a [jV] sequence to have a faster transition than a [iV] sequence. Liberman et al. (1956) suggest that this is the case, at least on perceptual grounds. In a perceptual experiment using simulated (drawn formant) speech stimuli, they find that the speed of transition from the formant starting point to the target formant state of the following vowel influences listeners' perceptions: fastest leads to a [CV] percept; slowest leads to a [VV] percept; in between leads to a [GV] percept. Studies of production have also confirmed temporal aspects regarding formant transitions to distinguish glides. In a study of English diphthongs Gay (1968) finds that the rate of transition of F2 serves as a reliable distinguisher between diphthongs.

A final timing factor to consider is the earliness of transition. That is, the distinction may not necessarily be about how fast the transition is, but how early it starts. The phonological interpretation that [j] is in the syllable margin with C_, while [i] is further structurally separate as a member of the syllable nucleus, would suggest a tighter gestural coordination between a glide and the preceding consonant, whether directly formalized (á la Gafos, 2002) or a result of the gestural planning of the relative segmental syllabifications. While the studies above, as well as Chitoran's (2002) acoustic analysis of Romanian diphthongs, examine transition speed, this study is intended to tease transition speed apart from transition earliness as potentially distinct characteristics. To summarize, multiple characteristics related to timing could therefore be central to distinguishing a [jV] sequence from a [iV] sequence: duration of the entire sequence, speed of transition into _V, or earliness of transition into _V. It could be a combination of these, or one might be a more consistent distinguishing factor.

It is important to note that all of the phonological accounts should predict the distinction to result in some kinds of timing differences like those presented above. No matter the featural representation, the glide will be part of the syllable margin, instead of the nucleus. Therefore, while factors related to timing may be found to convey this distinction, they do not necessarily rule out a place- or constriction-based representation. However, if we find *only* such timing characteristics to play a role, this may lend support to applying a syllabic pre-linking account and challenge the other representations proposed. But if timing differences are found in tandem with the other predictions presented above, this would still lend more support to those associated accounts than to concluding with a pre-linking account. This is because, if anything, a pre-linking (which we could also think of as the timing- or structure-only) account would predict the reverse in terms of formants and intensity. If the glide were identically featured yet forced to the syllable margin, the faster articulation might predict a lessened amplitude of articulator movements, resulting in formant centralization due to not fully reaching the target (Gay, 1981; Browman and Goldstein, 1990; Turner et al., 1995).

To summarize the acoustic predictions generated by these competing accounts: All accounts should predict some kind of timing difference due to syllabification, with a [jV] sequence showing a shorter overall duration, an earlier transition to [_V], and/or a faster transition. The place-based representation also predicts acoustic evidence that [j] is more front than [i], and that it therefore

has a higher F2. The constriction-based representation, instead, predicts that [j] has a lower acoustic intensity and/or a higher lingual articulation and therefore a lower F1. A finding of *only* timing differences, possibly in tandem with formant centralization, would lend stronger support to the syllabic pre-linking account.

3. Method

3.1. Elicitation

3.1.1. Participants

Nine speakers participated in this study. All were remunerated for their time. The study took about 40 minutes, including informed consent and a questionnaire eliciting demographic information. All were identified as native speakers of American English. All were white, non-Hispanic or Hispanic. Most were female (8/9). Their ages ranged from 19 to 37 years. A participant's particular US region they identified with was not tightly controlled for as this variable has not so far been shown to significantly differ across any regional varieties of American English. No speakers reported having ever been diagnosed with any speech- or hearing-related disorder.

3.1.2. Stimuli

The experiment elicited utterances of both pre-existing lexical items and nonce names. All were designed with the purpose of eliciting the [j]-[i] distinction in a C_V environment. This environment was chosen because it is one where the distinction does seem to be available, as discussed above. Each stimulus was embedded within a unique sentence that participants read aloud. This was done in two blocks: lexical items in the first, nonce names in the second. There were 8 LEXICAL ITEMS considered to carry this distinction in near-minimally paired environments. These are listed in Table 1, with the words categorized by their EXPECTED PRONUNCIATION and counterparts vertically paired. Word position, preceding segment, and stress placement was controlled across all words. As mentioned above, expected pronunciations were cross-checked, though admittedly subject to potential variation. At the time of writing, when looking across different sources' entries, an asymmetry was apparent for each pair in the expected direction, in agreement with Table

1. (For example, while Dictionary.com listed both pronunciations as options for *gardenia*, both Merriam-Webster and Cambridge listed only the [jə] pronunciation. On the other hand, the [iə] pronunciation of *Armenia* was the only option listed by the Cambridge Dictionary of American English, while Dicitonary.com and Merriam-Webster listed both options.) Furthermore, at the end of the experiment, participants were asked to complete a metalinguistic task of providing syllable counts, with each of the pre-existing lexical items of interest included. For each, the syllable count associable with the expected pronunciation was that more frequently provided. There were 9 additional lexical items elicited for future/followup analysis of this variable (e.g., *fjords*, *Tokyo*), but these did not have near-minimal pair counterparts and are not addressed further in this analysis.

Table 1: Stimuli: lexical items of interest

expected pronunciation				
[iV]:	Estonia	hernia	millennia	Armenia
[jV]:	pneumonia	California	Kenya	gardenia

The 17 lexical items were each assigned to a unique sentence. These assignments were kept constant. Attention was paid to placing the stimulus in a prosodically prominent position in the early half of the sentence. The following word in every sentence began with a voiceless labial obstruent, simply as a means of controlling the place and sonority of the consonant immediately following the sequence of interest. Two examples are provided in (2).

(2) a) The state of California passed a new bill.

b) The gardenia flower has a strong scent.

There were 40 NONCE NAMES designed to test if the distinction could be elicited productively. They were also designed with the intent of eliciting the distinction in a much more diverse array of phonological environments, to therefore examine what acoustic aspects convey it consistently. To elicit the distinction itself, pairs differing in ORTHOGRAPHY were made with the aim of eliciting [j] via the <y> grapheme and [i] via <i>. The preceding consonant and the position in the word were also manipulated. Three PLACES OF ARTICULATION of the preceding consonant were used: labial, coronal, and dorsal. Within each place of articulation, four MANNERS OF ARTICULATION were used: voiceless stop, voiced stop, voiceless fricative, and nasal (the latter two manners unavailable for the dorsal place in English's phonemic inventory). Finally, WORD POSI-

TION was also manipulated. Position here is defined in terms of the placement of the [Cj]/[Ci] sequence: initial vs. medial. Paired counterparts across the main condition of orthography were created, matched along the other three factors to therefore balance for potential phonological effects on the distribution of this distinction, as discussed above (§2.2). There were also 40 filler nonce names created, none incorporating the variable of interest.

Additional environmental factors within the nonce names were controlled. The following vowel (elicited via the <a> grapheme) was kept constant within each position: [ɑ́] for initial position and [ə] for medial position. For the initial-position stimulus pairs, the place of the following consonant was kept identical. For the non-target syllable in each stimulus, the inventory of nuclear vowels was [ɑ, i, u, o]. This provided some diversity while keeping nucleus weight constant. A final aspect of the stimuli was the use of acute accent <´> marks to represent stress placement. This was incorporated to keep participants from placing stress on the high front vocoid of interest (e.g., pronouncing *Súmia* as [su.mí.ə]). This factor also served as a distractor variable, with the 40 filler stimuli varying more unpredictably in stress placement (e.g., *Shóglubo*, *Blitú*, *Sóga*). Table 2 lists the nonce name stimuli along with their honorifics, which will be further explained next.

Table 2: Stimuli: nonce names (w/ honorifics)

		INITIAL		MEDIAL	
C_		<i>	<y>	<i>	<y>
LABIAL	/p_/	Dr. Piácho	Governor Pyásha	Coach Nópia	Officer Dápya
	/b_/	Mr. Biási	Mr. Byásu	Miss Shábia	Mrs. Chóbya
	/f_/	Mr. Fiáki	Officer Fyága	Mr. Gófia	Dr. Zúfya
	/m_/	Sister Miáshu	Professor Myáchi	Professor Súmia	Dr. Fímya
CORONAL	/t_/	Dr. Tiágu	Sister Tyáko	Governor Bítia	Mr. Pótya
	/d_/	Dr. Diáfa	Mr. Dyápu	Sister Módia	Sister Vádya
	/s_/	Officer Siáko	Professor Syági	Officer Kúsia	Officer Gísya
	/n_/	Sister Niáfi	Dr. Nyápa	Miss Vónia	Judge Búnya
DORSAL	/k_/	Professor Kiása	Mr. Kyáso	Mrs. Dókia	Mr. Púkya
	/g_/	Pastor Giáfu	Dr. Gyápi	Judge Nágia	Professor Tígya

Like the lexical items, each nonce stimulus was embedded in a unique sentence, with that sentence assignment remaining constant. Carrier sentences presented the nonce stimuli as surnames in sentence-initial position. The honorifics kept the stimuli away from completely phrase-initial position while still early in the sentence in a prosodically prominent position. All sentences were of the formula presented in (3) with some examples.

(3)

	HONORIFIC	+	STIMULUS	+	VERB	+	DIRECT OBJECT / ADJUNCT MODIFIER
a)	Mr.		Byásu		started		a band.
b)	Judge		Búnya		paints		beautifully.

Environmental factors within the sentences were also controlled. For medial-position stimuli, the onset segment of the following word in the sentence was always a voiceless labial obstruent (the same method for controlling the following segment as that used for the real word stimuli described above). For initial-position stimuli, the preceding honorific was always [ɪ]-final.

3.1.3. Procedure

The study took place in the Phonetics and Experimental Phonology laboratory at New York University. Participants were seated in a sound-attenuated booth at a desk with a computer screen in front of them. Their speech was recorded with a Shure SM35-XLR head-mounted microphone connected to a Marantz PMD 660 audio recorder (44.1kHz sampling). Sentences were presented on the computer screen one at a time. The participant would read the sentence aloud and advance to the next by pressing the down arrow on a standard keyboard. This method expressly avoided auditory repetition, so that no such distinction nor its implementation could be auditorily primed. Previous research suggests that speakers' productions can be phonetically influenced by previous exposure (e.g., Goldinger, 1998; Namy et al., 2002; Fowler et al., 2003; Gentilucci and Bernardis, 2007). Therefore, elicitation was only done by orthographic presentation, so that participants' productions could not be influenced by previous exposure in any way. The researcher was present in the sound booth to provide training for the nonce stimuli section and ask the participant to repeat any stimulus if needed. All participants were led through the same procedure with the same stimuli and presentation described below.

The first block consisted of the 17 sentences containing pre-existing lexical items. Sentences were randomized, then near-minimal pairs were moved to allow substantial space between the counterparts. This was repeated to result in 4 cycles through the stimuli, with the spacing of near-minimal pair counterparts across cycle boundaries also manually adjusted.

The second block consisted of the 40 sentences containing nonce names (and 40 with filler nonce names). First, the nonce names of interest (targets) were randomized. Then, ordering was

adjusted to put maximum distance between near-minimal pair counterparts: those matching in environmental factors and differing in <y> vs. <i> orthography. Then, the filler stimuli were randomly ordered and added, one after each target stimulus, so that the cycle would alternate between target and filler stimuli. This was repeated to result in 4 cycles through the stimuli. After this, spacing of near-minimal pair counterparts was given similar attention across cycle boundaries.

Between the two blocks, there was a short training session regarding the nonce stimuli. The researcher told participants that they would be encountering unfamiliar last names. They were told that the names use only 4 vowels—[ɑ], [i], [u], and [o]. They were instructed to be consistent in pronunciation, thinking one letter equals one sound: e.g., the letter <g> should always be pronounced as [g], and never as [d͡ʒ]. They were then instructed that the vowel marked with an acute accent <´> was stressed.

After this instruction, three cycles of training stimuli were presented. Training stimuli were all made according to the filler stimulus formula. None included <y> or any <iV> sequence; some included simplex <w> onsets. The first cycle was auditory and orthographic repetition. There were ten training stimuli consisting of just an honorific + name. A pre-recording of the stimulus uttered by another English speaker played automatically with each slide showing the orthography, and the participant would repeat it. In these pre-recorded utterances, when an <a> was final and not stressed it was reduced to a schwa, which participants followed naturally. The second cycle removed the auditory component. There were five training stimuli consisting of just an honorific + name. In this cycle, a pre-recorded utterance was no longer played and the participant would read the orthographically presented stimulus aloud. The researcher provided feedback after any errors (which were usually regarding stress placement). The last training cycle consisted of four stimuli in full sentence form. Participants were told that they would now be reading complete sentences with these names. They were told that it was important to not pause within a sentence and that they may be asked to repeat if they paused within. However, they were informed that there was no time limit and that they may say the sentence in their head before saying it out loud.

Feedback was provided during the second block. However, no feedback was ever given regarding the variable of interest. The researcher did nothing if, on a <y> stimulus, the par-

ticipant's utterance was perceived as a [iV] sequence or, for a <i> stimulus, the participant's utterance was perceived as a [jV] sequence. Both of these behaviors were perceived to occur, though, suggesting that phonological effects on the distribution did sometimes override the orthographic elicitation. No participant was perceived to categorically produce only [jV] or [iV] across the orthographic presentations. One phenomenon that did elicit feedback regarding <y> and <i> was the pronunciation of either as the [aɪ] diphthong. This was not common, but did occur a few times with more than one participant. In such cases, the feedback was framed along the following lines: "Don't pronounce the letter <i> or <y> as [aɪ]. The only vowels are [ɑ], [i], [u], and [o]." Feedback never included an utterance of a [jV] or [iV] sequence by the researcher. Common errors eliciting feedback were misplacement of stress, pausing, and segmental errors not within but sometimes neighboring the sequence of interest.

3.2. Analysis and predictions

There were 1672 utterances examined, after excluding tokens which were produced in an unexpected way (e.g., the vocoid of interest pronounced as [aɪ], a relevant neighboring segment mispronounced, stress misplaced, the sequence held out as a speech delay). Praat software (Boersma and Weenink, 2015) was used for segmentation and analysis. The entire [jV]/[iV]-expectant vocalic sequence was segmented. The beginning (henceforth vocalic onset), was identified as the onset of F1 after an obstruent or re-strengthening of formants after the release of a preceding nasal. The end (henceforth vocalic offset) was identified as the severe reduction in amplitude or complexity of the formants attributable to the following consonant. All acoustic measurements were performed over this entire vocalic sequence.

The following reviews what this distinction might look like acoustically. In line with a place-based representation, we would expect that [j] has more anterior raising of tongue mass than [i], and therefore a higher F2. In line with a constriction-based representation, we would expect that [j] has a higher lingual articulation and tighter constriction than [i], and therefore a lower F1 and lower acoustic intensity. As predicted by all accounts (now including that of syllabic pre-linking), timing may also play a role, with a [jV] sequence being shorter overall, having an earlier transition, and/or having a faster transition than a [iV] sequence.

Figure 1 shows spectrograms of utterances by the same speaker of *Estonia* [CiV] and *pneumonia* [CjV] appearing to exhibit the distinction. Inspection confirms some of the predictions above. The [jV] sequence is of a shorter duration overall than that of [iV]. In terms of the F2 trajectory, the [iV] sequence appears to take longer to reach its maximum before transitioning downward toward the following vowel, suggesting that the [jV] transition starts earlier. However, the F2 max is greater for [iV] than for [jV], suggesting that [j] is *less* front than [i]—the opposite of that predicted by the place-based representation. The intensity tracker also shows a greater jump in the [jV] case (+6.1dB) than the [iV] case (+3.5dB), suggesting that [j] has a lower intensity relative to _V than [i] does. And [jV] has a lower F1 min, suggesting a higher lingual articulation. Other than that of F2 max, these observations are all in line with the acoustic phonetic predictions discussed above.

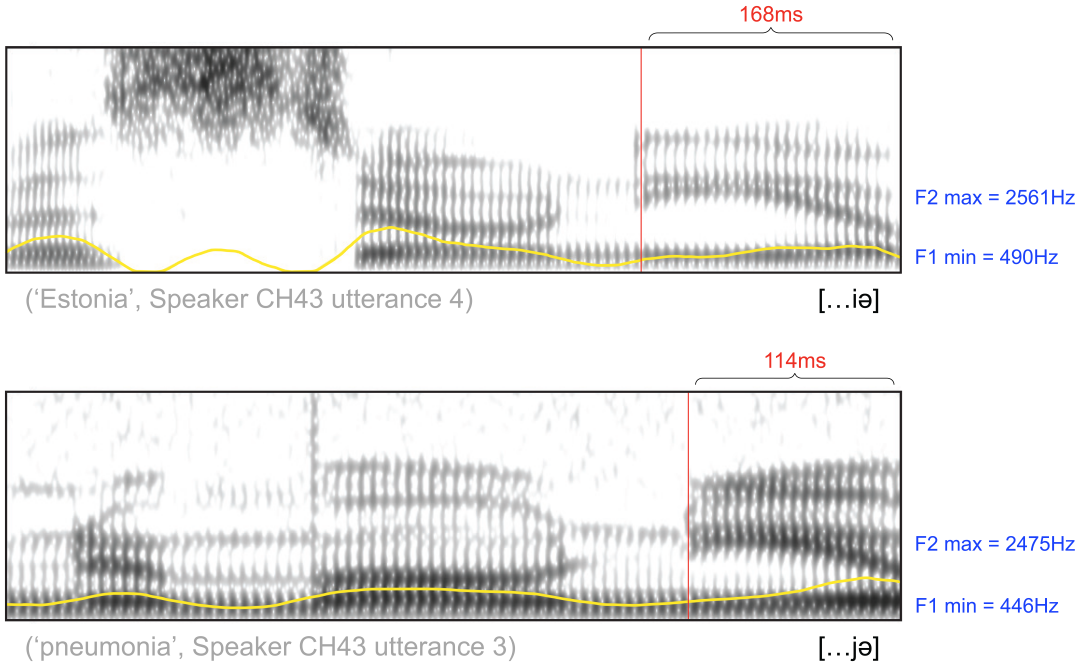

Figure 1: Example utterance spectrograms

Spectrograms of utterances by the same speaker of a near-minimal pair expecting and appearing to exhibit the distinction of interest. The vertical red line shows the vocalic onset and the end of the spectrogram is where the vocalic offset was segmented, with the duration of the entire vocalic sequence noted. The yellow line is Praat's intensity tracker. The F2 max and F1 min of each vocalic sequnce of interest are also noted.

Table 3 summarizes the measurements, predictions, and how each relates to the competing representation accounts. All measurements were performed across the entire segmented vocalic sequence using a Praat script. The maximum value of F2 was identified and recorded (F2 max)

to examine frontness: a higher value for [j] would mean that it is more front than [i], therefore supporting a place-based representation. The minimum value of F1 was identified and recorded (F1 min) to examine height of lingual articulation: a lower value for [j] would mean that it is of a higher lingual articulation than [i], therefore supporting a constriction-based representation. The minimum intensity value was recorded and subtracted from the maximum intensity value (intensity range): a greater intensity range for [jV] would mean that [j] has a lower intensity relative to _V, therefore supporting a constriction-based representation. The amount of time between the voicing onset and the timepoint at which F2 max occurred was also recorded (F2 max time) to examine the transition starting point (following Chitoran, 2002 and Ren, 1986), predicting a [jV] sequence's transition to begin earlier. (See §4.3 for further discussion regarding the choice to treat this measurement of earliness as absolute—milliseconds from voicing onset—rather than relative—such as percentage of the entire vocalic sequence's duration.) The timepoint and value of the F2 minimum were recorded and used to calculate the slope of F2's transition between that point and F2 max (F2 slope), predicting a [jV] sequence to have a greater F2 slope and therefore a faster transition (as suggested by Liberman et al., 1956). Finally, the overall duration was measured between vocalic onset and offset, predicting a [jV] sequence to have an overall shorter duration than a [iV] sequence.

Table 3: Measurements and competing acoustic predictions

MEASUREMENT	PREDICTION	REASON	ACCOUNT
F2 max	[iV] < [jV]	[j] more front than [i]	place
F1 min	[iV] > [jV]	[j] lingual articulation higher than [i]	constriction
intensity range	[iV] < [jV]	[j] more constricted than [i]	
F2 max time	[iV] > [jV]	[jV] has earlier transition than [iV]	all accounts
F2 slope	[iV] < [jV]	[jV] has faster transition than [iV]	
duration	[iV] > [jV]	[jV] = 1σ; [iV] = 2σ	

Of course, one possibility is that there is no significant difference along any measurement, which would not support the hypothesis that a distinction was elicited (at least as detectable by the measurements taken here). An observation in the *opposite* direction of one of the predictions above would motivate *ruling out* the associated account. For example, an observation that [j] has a *lower* F2 max would suggest that a place-based approach is not a good fit for representing the distinction at hand. An observation that any predictions from the first two sets are borne out would

lend support to the associated account(s). For example, an observation that [j] has a *higher* F2 max would support applying a place-based representation. However, that observation in tandem with a lower F1 min or wider intensity range could be interpreted as inconclusive between place- and constriction-based accounts, or as support for a hybrid approach that both features are part of this distinction's specification (e.g., Nevins and Chitoran, 2008). Furthermore, as discussed above (§2.3), if we observe any predictions from the first two sets in tandem with any from the final set of timing-related predictions, this would still lend more support to those accounts than to a syllabic pre-linking account, since a difference in syllabification and therefore timing is predicted by all accounts. Strongest support for a syllabic pre-linking account would come from observing *only* timing-related differences.

4. Results

In this section, all of the acoustic measurements (previously summarized in Table 3) are submitted to statistical tests to examine whether they exhibited significant differences across the near-minimal pairs, which would suggest the distinction was successfully elicited. Utterances are only coded for expected output, not perceived production. For pre-existing lexical items (§4.1), this is the factor of expected pronunciation (as previously given in Table 1). For nonce stimuli (§4.2), this is the factor of orthography, which expects a <y> → [j], <i> → [i] mapping. Any significant result would suggest that the distinction was successfully elicited, with paired counterparts distinguishable along any acoustic measurement(s) with a significant effect.

4.1. Pre-existing lexical items

A linear mixed-effects model was performed for each of the acoustic measurements using the lme() function from the nlme package (Pinheiro et al., 2017) for the R statistical programming environment (RCoreTeam, 2015). This tested expected pronunciation as the independent variable for its effect on each acoustic measurement as the dependent variable. To capture the paired nature of the experimental design, a random effect (intercept and slope) was included for each combination of speaker and near-minimal pair. For example, anatomical differences might lead formants to have a higher average for one speaker than another, and that could also affect the

magnitude of the difference in any formant used by that speaker to convey the distinction. Or, one speaker might not exhibit the distinction for one word pair, though they exhibit it for another. Or, timing-related factors might be used differently for one word pair than another, such as the distinction being realized differently between the *Kenya* + *millennia* pair and the *California* + *hernia* pair due to differing syllable counts, seeing as entire word length can effect the duration of syllables within the word (Lindblom, 1968).

The results of this analysis are presented in Table 4. The first three columns are descriptive statistics. 'Percent [i] > [j]' describes what proportion of near-minimal utterance pairs exhibit a difference in the [i] > [j] direction; the farther this number is from 50% means an acoustic difference is less chance-like in its patterning and therefore more consistent in conveying this distinction. The means of both categories are also provided. The final two columns are results of the statistical models applied. The 'Coefficient: [j]-expectant' is how much and in what direction the model predicts an acoustic measure to differ from the [i]-expectant prediction (the intercept) when the utterance is of a [j]-expectant stimulus.

Table 4: Results: pre-existing lexical items

Measurement	Percent [i] > [j]	Mean: [i]-expectant	Mean: [j]-expectant	Coefficient: [j]-expectant	*p*
F2 max time	94%	35.67ms	19.43ms	-16.49ms	9.99*e*-16 ***
duration	83%	167.03ms	130.23ms	-37.09ms	2.49*e*-09 ***
F2 max	75%	2609Hz	2543Hz	-67Hz	0.00066 ***
F2 slope	25%	10.078Hz/ms	10.791Hz/ms	+0.769Hz/ms	0.04965 *
intensity range	31%	5.57dB	6.57dB	+0.992dB	0.03172 *
F1 min	56%	431Hz	423Hz	-6Hz	0.44118

Descriptive statistics and results of linear mixed-effects modeling per measurement across the factor of expected pronunciation. Measurements are ordered by their consistency of conveying the distinction: how far, in either direction, Percent [i] > [j] is from 50%.

The expected glide-vowel distinction across the near-minimal word pairs appears borne out in the data along multiple acoustic dimensions. The [j]-expectant counterparts have significantly earlier transitions into the following vowel, as represented by the earliness of F2 max, and significantly shorter durations of the entire vocalic sequence. They also have significantly wider intensity ranges, suggesting that [j] has a lower intensity relative to that of the following vowel. A difference in frontness, as represented by F2 max, is also significant but in the *opposite* direction than that predicted by a place-based representation: [j] has a significantly lower F2 max, and

thus a less anterior constriction than [i]. And the measurement of F2 slope suggests that [j] has a significantly faster transition into the following vowel. F1 min shows no significant effect. Figure 2 provides visualizations of the factors found to significantly exhibit the distinction.

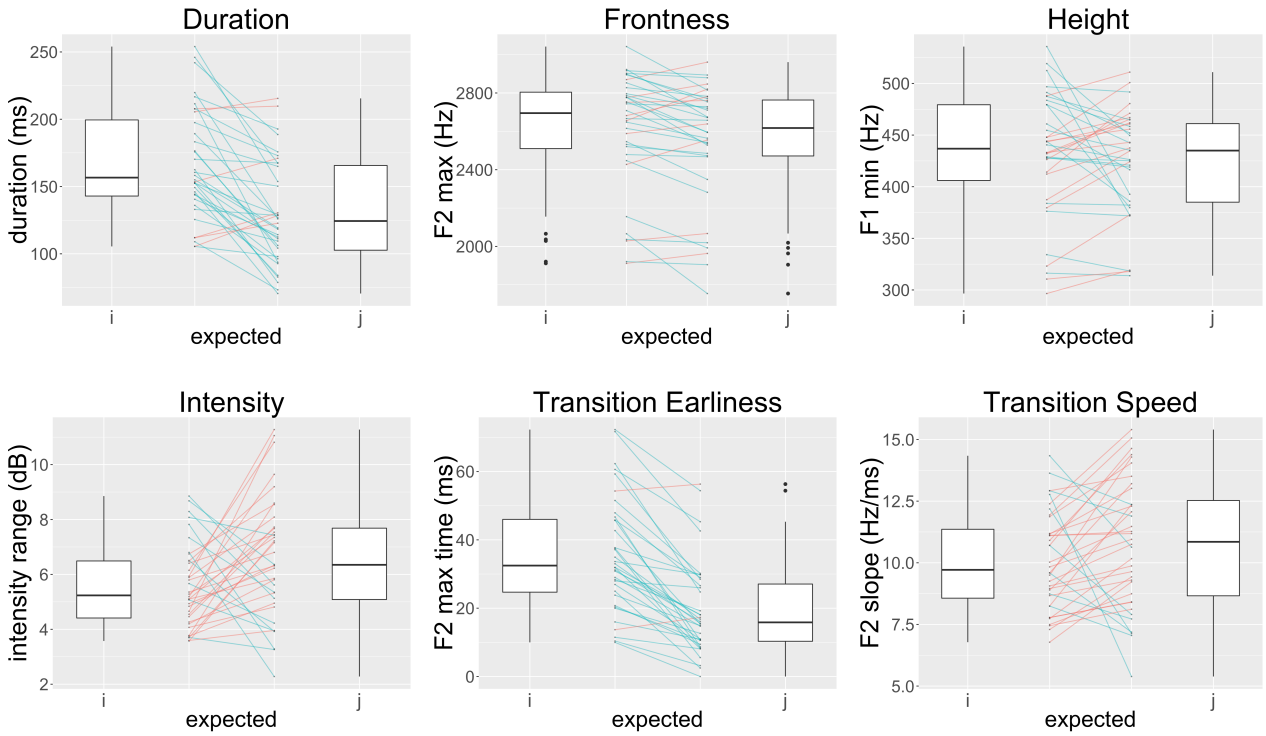

Figure 2: Pre-existing lexical items

Visualizations provide box plots across the two expected pronunciation conditions. Points represent the measurement for each speaker's utterance of each word (averaged across repetitions and grouped by expected pronunciation alongside the respective box plot). Lines connect each pair's counterparts, with a green line representing that [i] > [j] and a red line representing that [i] < [j].

These results suggest that a [j]-[i] distinction is present between near-minimal pre-existing word pairs and that timing may be the most reliable distinguisher between [j] and [i]: [jV] sequences are shorter than [iV] sequences, seemingly brought about by an earlier and faster transition into [_V]. The results also challenge applying a place-based representation, with a difference in frontness found to be significant but in the reverse direction of that predicted by this representation: [j] is *less* anterior than [i]. This could be a by-product of timing, as discussed earlier (§2.3): [j] being faster than [i] could result in more reduction and formant centralization, therefore keeping [j] from reaching its front target (however anterior that target may be) and in this case resulting in a significantly less anterior realization than that of [i]. Results speak in support of applying a constriction-based representation. The intensity range effect suggests that [j] has a lower intensity, relative to that of the following vowel. Furthermore, the lack of a significant

reverse effect for F1 min like that observed for F2 max could suggest that the target is actually a higher articulation than [i] but production is reduced due to the faster articulation, though not to a realization lower than (only insignificantly different from) that of [i]. In summary, these results challenge applying a place-based representation and support applying a constriction-based representation, at least regarding the production of pre-existing near-minimal word pairs in American English.

4.2. Nonce stimuli

This section extends the same analysis to the nonce stimulus data, with analogous linear mixed-effects models of each measurement across the condition of stimulus orthography: with a <i> orthography expecting a [i] output and a <y> orthography expecting a [j] output. Again, a random effect was specified for each combination of speaker and near-minimal pair. The results are presented in Table 5, with the format exactly like that of Table 1 regarding the pre-existing lexical item results.

Table 5: Results: nonce stimuli

Measurement	Percent [i] > [j]	Mean: [i]-expectant	Mean: [j]-expectant	Coefficient: [j]-expectant	*p*
duration	66%	199.89ms	186.39ms	-13.21ms	6.44*e*-08 ***
F2 max time	66%	23.93ms	19.59ms	-4.38ms	3.76*e*-06 ***
intensity range	65%	10.06dB	9.28dB	-0.763dB	0.00014 ***
F2 max	59%	2632Hz	2619Hz	-12Hz	0.08380 ·
F2 slope	45%	9.249Hz/ms	9.563Hz/ms	-0.278Hz/ms	0.11019
F1 min	51%	371Hz	368Hz	-3Hz	0.36107

Descriptive statistics and results of linear mixed-effects modeling per measurement across the factor of stimulus orthography. Measurements are ordered by their consistency of conveying the distinction: how far, in either direction, Percent [i] > [j] is from 50%.

Again, the expected glide-vowel distinction across the near-minimal nonce stimulus pairs appears borne out in the data along multiple acoustic dimensions. The [j]-expectant counterparts have significantly earlier transitions into the following vowel and significantly shorter durations of the entire vocalic sequence. There is again a significant effect on the intensity range, however this is in the reverse direction: [i]-expectant stimulus utterances show a greater intensity range across the vocalic sequence than [j]-expectant stimulus utterances. It's possible that this is a task effect. Recall that stress placement was explicitly marked in the nonce stimuli and used as a distractor

variable during the experiment. Subjects may have been hyperarticulating stress by using a wider than normal intensity range to distinguish stressed syllables from unstressed syllables. The hyper-differentiation of intensity between syllables may be overriding any observably lower intensity of [j]. However, this potential for reversal does suggest that intensity may not be the most reliable characteristic of this distinction. The remaining measurements pattern in parallel with the results of the pre-existing lexical items discussed above. F2 max again patterns counter to what would be predicted by a place-based representation, with [j] having a lower F2 max (this time approaching, while not reaching, significance) and therefore a less anterior articulation. Figure 3 provides visualizations of the factors found to significantly exhibit the distinction.

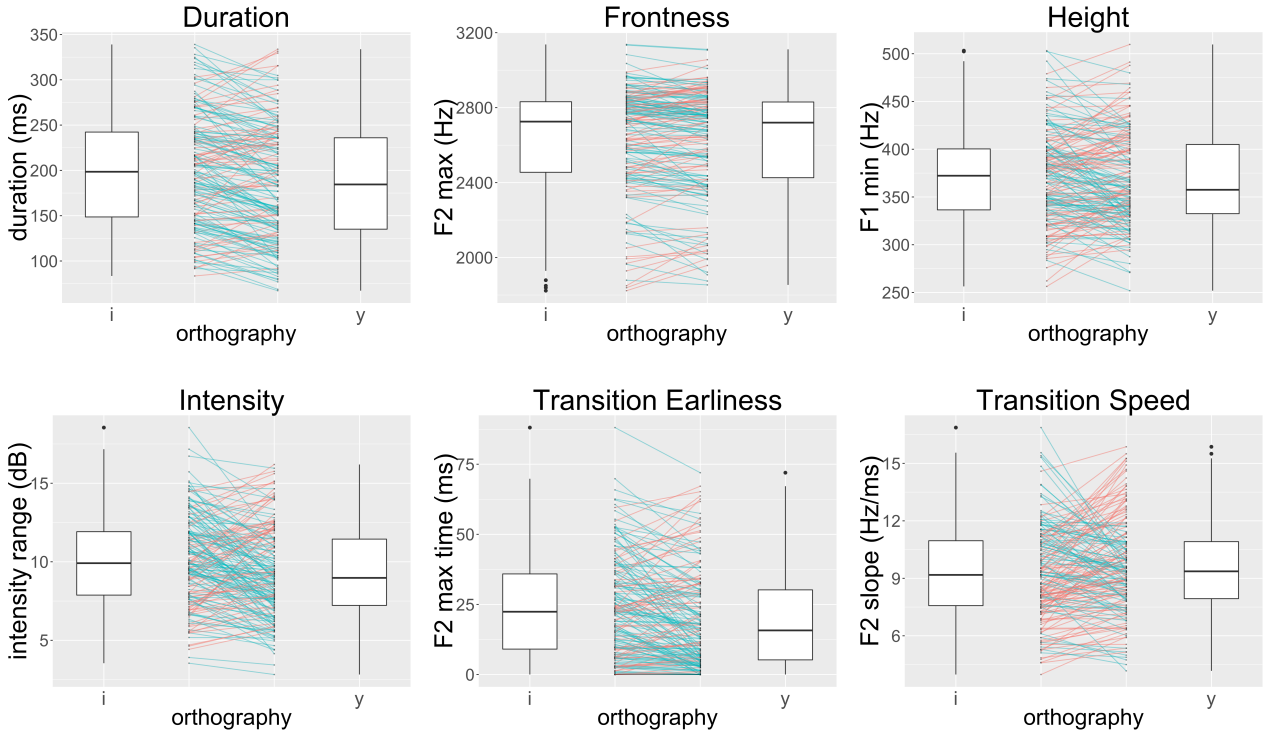

Figure 3: Nonce stimuli

Visualizations provide box plots across the two orthography conditions. Points represent the measurement for each speaker's utterance of each word (averaged across repetitions and grouped by orthography alongside the respective box plot). Lines connect each pair's counterparts, with a green line representing that [i] > [j] and a red line representing that [i] < [j].

These results suggest that the [j]-[i] distinction observed between pre-existing near-minimally paired lexical items (§4.1) is also productively extended to new words, as elicited via the <y> vs. <i> orthographic distinction, and across a wider variety of surrounding environments. They further suggest that transition earliness and overall vocalic sequence duration are the more consistent acoustic dimensions that convey this distinction: [jV] sequences are shorter than [iV]

sequences, with the transition into [_V] coming earlier after [j] than after [i]. The nonce stimulus results continue to challenge applying a place-based representation to this case, with [j] again found to have a lower F2 max and therefore a less anterior articulation—the reverse of that predicted by this representation. However, these results speak less strongly in favor of a constriction-based representation, with [j] now appearing to have a greater intensity with respect to that of the following vowel. How to conclude or proceed based on these observations will be further discussed below (§5).

While the central pursuit of the nonce stimulus part of this study is to examine what acoustic characteristics consistently convey this distinction across a more diversified array of surrounding environments, the data may also speak to how those environmental factors constrain the distinction's availability. Table 6 reports measurements of what appears to be the most consistent characteristic, transition earliness as measured by F2 max time, across the different environmental conditions of word position and C_ place that were manipulated in the nonce stimuli. Though linear regression modeling of each environmental condition's effect (testing for a significant interaction between condition and orthography as a predictor of F2 max time) finds no effect to be significant, the distinction's availability does appear to pattern in some expected ways by being realized more consistently in certain conditions. It is apparently more available when the preceding consonant is in medial position rather than initial position. This, as previously discussed above (§2.2), could be due to consonants that would otherwise disprefer sharing a complex onset with the glide being more readily parsed as the coda of the preceding syllable when the consonant is not word-initial. The distinction is also apparently more available when the preceding consonant is coronal or labial and less available when the preceding consonant is dorsal (the strongest case appearing to limit the distinction's availability). This suggests a homorganicity constraint banning [Cj] sequences when the preceding consonant is dorsal (to be discussed further still in §5).

Table 6: F2 max time across nonce stimulus environmental conditions

Factor	Condition	Percent [i] > [j]	Mean: [i]-expectant	Mean: [j]-expectant	Mean of Differences
position	medial	71%	22.36ms	16.62ms	5.75ms
	initial	62%	25.47ms	22.53ms	2.94ms
C_ place	coronal	71%	26.75ms	21.18ms	5.58ms
	labial	69%	27.37ms	22.76ms	4.61ms
	dorsal	53%	11.47ms	10.17ms	1.3ms

Descriptive statistics of F2 max time (representing transition earliness) across manipulated conditions of the surrounding environment. Within each factor, conditions are ordered by how consistently the measurement of F2 max time exhibits the distinction: how far Percent [i] > [j] is from 50%.

4.3. Transition earliness: absolute vs. relative

Given that transition earliness appears to be the most consistent differentiating characteristic of this distinction, it is briefly given some more nuanced attention here. In the above analyses, transition earliness is treated as an absolute measurement: How many milliseconds after the onset of a [jV]/[iV] sequence is the maximum F2 reached before its descending transition into the following vowel begins? However, we know that the duration of a segment can be influenced by segment-extrinsic factors like speech rate (Goldman-Eisler, 1968; Grosjean and Lane, 1976; Gay, 1978; Hirata, 2004), whether a segment appears in a stressed syllable (Lindblom, 1963; Klatt, 1975), and the crowding of its prosodic environment, such as the number of syllables in a stress group (Crystal and House, 1990). This consideration might, therefore, motivate examining transition earliness as a relative, proportional measurement rather than an absolute measurement: How far percentage-wise into a [jV]/[iV] sequence (no matter its entire duration) does the transition to [_V] begin? A [jV] sequence may show an average F2 max occurrence time of 20ms and an average total duration of 185ms, but it might be hypothesized to come later than 20ms after the onset in a longer [jV] sequence of, say, 200ms total. Therefore, a relativized measurement may capture the distinction better across a diverse array of environments by accounting for this potential variation.

On the other hand, effects on segmental duration are not entirely absolute or consistent. Studies examining the effects of speech rate have observed that pauses (Goldman-Eisler, 1968) and vowels (Kozchevnikov and Chistovich, 1965; Gay, 1978) are the main loci of duration changes across different speech rates. Furthermore, Klatt (1973) argues that segments may have intrinsic

absolute durations that become apparent when considering segment-extrinsic influences. When testing a model of segment duration incorporating multiple factors that have been found to influence it, Klatt observes that vowel categories seem to have respective floors of compressibility at which point the model's predictions of shorter durations fail. Absolute duration may therefore play an important role in the characterization of a glide as well. Like Klatt observes regarding vowels, there could be a floor of compressibility for glides, albeit shorter. A duration shorter than that floor might, instead, resemble the very fast formant transition after a consonantal constriction. This is supported at least on perceptual grounds by Ohala's (1978) analysis that a change in Southern Bantu in which palatalized labial stops /p^{j}/ changed to coronal stops /t/ is due to the reanalysis of [^{j}] as a formant transition. (The coronal, as opposed to dorsal end result may be explained by the labial [p]: the high starting point of the F2 transition that would result from palatalization, and therefore resemble a dorsal transition, could have been mitigated by the low starting point after the labial constriction and therefore result in something in between the two.) We might also imagine a ceiling of expandability at which point a glide is no longer glide-like but vowel-like, no matter how long surrounding segments may be. This double-sided bounding of the duration of a glide is also at least perceptually supported by Liberman et al.'s (1956) finding that when increasing the speed of formant transition from a high F2 starting point, listeners' percepts change from [iV] to [jV] and then to [gV].

The data below provide a fuller description of transition earliness. In Table 7, the timepoint at which F2 max occurs is represented both in Relative (percentage of entire vocalic sequence duration) and Absolute (milliseconds after vocalic onset) terms. Figure 4 provides a series of Smoothing Spline ANOVA plots (Gu, 2002) of the F2 contours of two pre-existing lexical item pairs: one pair of those most similar in the entire form (*Estonia* + *pneumonia*), and another pair less so (*millennia* + *Kenya*). These plots are based off 50 equidistant measurements of F2 across the entire vocalic sequence, fitting curves to the datasets being compared and providing Bayesian confidence intervals to determine areas of significant difference between the formant contours.[3] The Relative versions show the F2 contour in terms proportional to the duration of the entire vocalic sequence, where the x-axis is the ordinal number for each of the 50 equidistant measurements ('timepoint'). For the Absolute versions, the x-axis is the conversion of these measurement points to their absolute duration ('raw timepoint': milliseconds after vocalic sequence onset).

Table 7: Results: transition earliness (absolute vs. relative)

	Percent [i] > [j]	Mean: [i]-expectant	Mean: [j]-expectant	Coefficient: [j]-expectant	*p*
Pre-existing					
absolute	94%	35.67ms	19.43ms	-16.49ms	9.99*e*-16 ***
relative	89%	20.9%	14.6%	-6.3%	4.69*e*-08 ***
Nonce					
absolute	66%	23.93ms	19.59ms	-4.38ms	3.76*e*-06 ***
relative	61%	11.4%	9.8%	-1.6%	.00073 ***

Descriptive statistics and results of linear mixed-effects modeling per measurement across the factor of expected pronunciation.

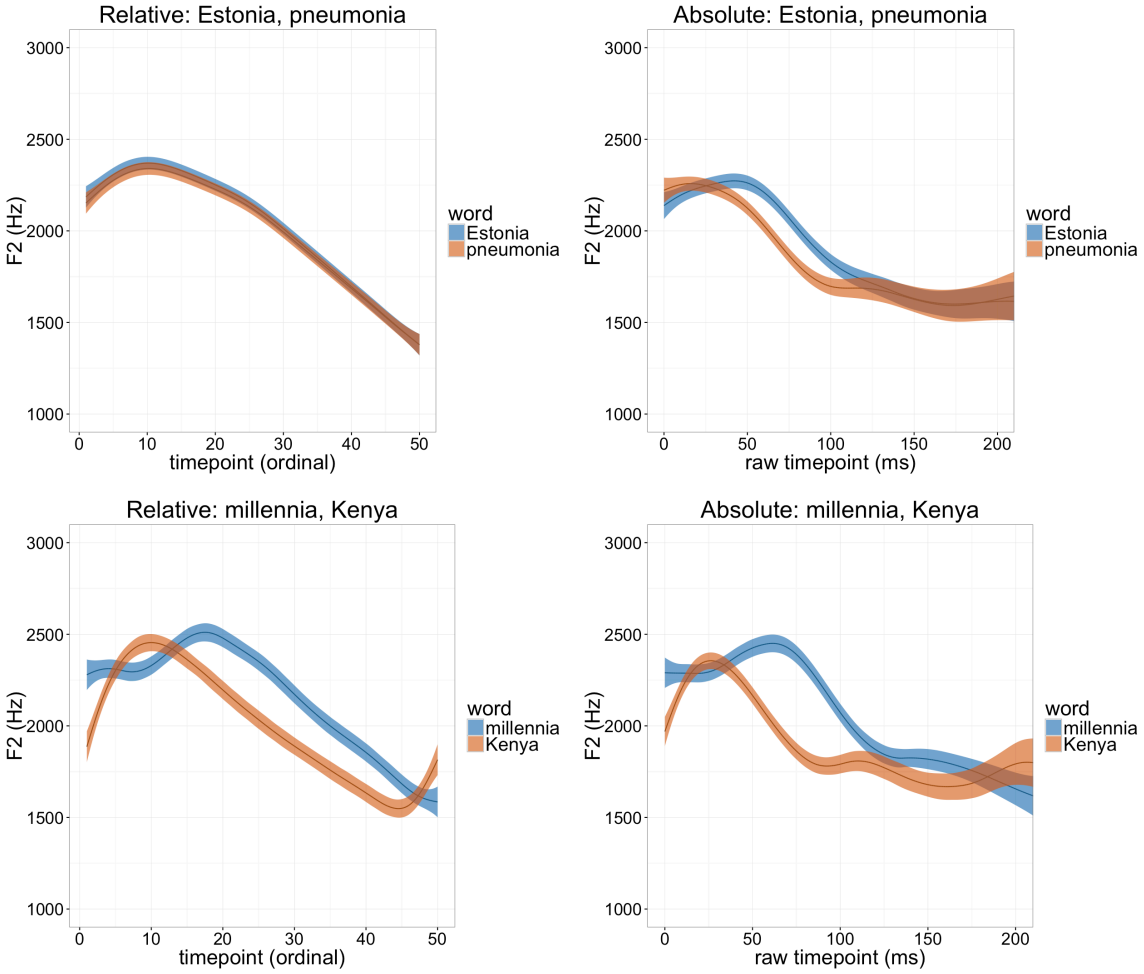

Figure 4: Smoothing Spline ANOVA plots

Plots on the lefthand side represent the x-axis in Relative terms, with 50 timepoints (evenly spaced across the entire vocalic sequence duration) expressed ordinally. Plots on the righthand side are in Absolute terms, with the x-axis converted to the amount of time (ms) between each point and the onset of the utterance's entire vocalic sequence. In each plot, the earlier half is that of primary interest containing the high front vocoid; the latter half represents the following vowel and also some indication of the transition into the initial segment of the following word.

These results demonstrate that both the absolute and relative approaches to the measure-

ment of transition earliness significantly reveal the distinction. However, they also suggest that an absolute approach to measuring transition earliness may be more consistent at capturing it. The results in Table 7 suggest that looking at how many milliseconds after the vocalic sequence's onset the F2 max is reached is a more consistent identifier of this distinction than looking at how far proportionally into the entire vocalic sequence's duration it's reached. And a more holistic analysis of the F2 contours in Figure 4 suggests that the contours are not always identified as significantly distinct when relativized to the entire sequence's duration, rather than anchored to real time. As discussed above, the relative approach is meant to account for other effects, such as a less exact environmental pairing. However, the absolute approach does seem to still capture the distinction across cases and fares better at doing so in a case of more similar environmental pairing.

Furthermore, when examining the Absolute plots, it is apparent that the confidence intervals become wider toward the end of the contour (the right side). This is due to variation in the duration of the entire sequence: when some utterances have shorter durations, there become fewer measurement points that can be referred to in the calculation of the confidence interval. So the entire sequence duration seems to vary, but the [j]-[i] distinction is still apparent when examined in absolute terms. The combination of these observations suggests that this variation of the entire sequence's duration may be more attributable to varying duration of the following vowel (corroborating findings mentioned above: e.g., Kozchevnikov and Chistovich, 1965; Gay, 1978). This somewhat orthogonal behavior lends support to the notion that, in [jV] and [iV] sequences, the prior and latter segments are truly separate concatenated segments rather than parts of a single complex unit.

5. Discussion and Conclusions

The results of this study suggest that there is a distinction between the [j] glide and [i] vowel available to native speakers of American English. This is elicited in utterances of near-minimally paired lexical items. It is also extended productively to nonce stimuli, elicited solely by orthography. Analysis identifies what acoustic aspects play a consistent role in the production of this distinction. The glide appears to most consistently be characterized by an earlier transition to

the following vowel and, likely as a result, a shorter overall duration of the vocalic sequence. Results from the pre-existing word pairs also suggest the glide to have a lower acoustic intensity, though this effect was reversed in the nonce stimulus production task (possibly as a task effect due to increased focus on stress placement). And while [j] is not shown to have a significantly higher and tighter lingual articulation (i.e., there is an insignificant difference in F1 min), neither is it shown to have a significantly lower and more open articulation. On the other hand, acoustic measurements do show a significant difference in articulatory frontness (measured by F2 max), suggesting that [j] is significantly less anterior (with a lower F2 max) than [i].

This further understanding of the acoustic character of this distinction serves us in multiple ways. It documents the distinction and aids in future approaches to identifying and segmenting it, which may help improve and increase its future documentation and allow for more robust analysis of its distribution and variability. The acoustic characterization can also contribute to the choice between different phonological representations considered. The finding that [j] has a significantly less anterior articulation supports ruling out a place-based representation (such as that proposed by Levi 2004; 2008, or a hybrid of it like that proposed by Nevins and Chitoran, 2008), which would have predicted [j] to be significantly *more* anterior than [i]. While the timing-related aspect of transition earliness was identified as the most consistent, this does not necessarily support applying a pre-linking account and ruling out a constriction-based representation, as all accounts considered generate such a timing-related difference as a by-product of syllabification. The observation that [j] is significantly less anterior could be attributable to reduction and centralization caused by the faster articulation of the glide. This could motivate us to update our predictions regarding the other formant measure of F1, leading us to consider only the stronger threshold of a parallel reverse effect (a significantly higher F1 min and therefore lower, more open lingual articulation) as support for ruling out a constriction-based representation and concluding with a pre-linking account. That is, if the significantly less anterior articulation of [j] is attributable to reduction due to its faster articulation, the lack of a significantly lower articulation could suggest a higher underlying target (or at least stronger resistance to reduction along the dimension of constriction height than anteriority). And this is coupled with the fact that, at least in the pre-existing word pairs, [j] is found to have a lower acoustic intensity. The results of this study are therefore still consistent with a constriction-based representation of the distinction.

Furthermore, characterizing the acoustics of this distinction may further our understanding of its phonological distribution. As discussed at the beginning of this paper (§2.2), and suggested by the results (Table 6), aspects of the surrounding environment appear to constrain the availability of this distinction. For example, the place of articulation of the preceding consonant appears to constrain the glide's appearance: [j] seems dispreferred when following a dorsal consonant. This constraint on the distribution of this distinction may be explained as a result of the distinction's acoustic realization and its apparent reliance on transition earliness. As discussed in §4.3, while the distinction requires [j] to have an earlier transition to the following vowel, if it is too early and fast there is potential for [j] to be misperceived as the formant transition cue in a [CV] sequence with a dorsal consonant. And this explanation would also hold for the apparent dispreference of [w] after labial constrictions (Clements and Keyser, 1983).

There are multiple further directions of inquiry that this study motivates. One is to examine the perception of this distinction, both in terms of cueing and contrast. The analysis above examines acoustic measurements as characteristics of this distinction: details of the acoustic signal that exhibit significant differences in production. Some characteristics (e.g., transition earliness and duration) appear more consistent and reliable than others (e.g., intensity). It would be helpful to know if this characterization of the distinction's *production* is a reasonable representation of how it is cued to the *perception* of the human listener. Do the same characteristics play the same roles as cues in the listener's perceptual distinction of glides from vowels? A perception experiment cross-manipulating these acoustic dimensions and testing them as predictors of participants' responses could provide beneficial comparison to the observations made here. A finding that intensity and F1 play a strong role in perceptually cueing the distinction could further strengthen our confidence in selecting a constriction-based approach as the optimal representation. Furthermore, manipulating the duration of surrounding segments to test for boundary shifts of this distinction (e.g., Ainsworth, 1974; Hirata, 1990) could speak further to the question of how absolute or relative the cue of transition earliness is (§4.3): e.g., Is there a duration beyond which a high front vocoid is categorically considered a vowel rather than a glide, irrespective of how long the following vowel may be? Also, while this study suggests a *distinction* that can be produced by speakers of American English, it does not necessarily demonstrate a *contrastive* function of it in the grammar. That is, none of the lexical item or nonce stimulus pairs tested here are exact minimal pairs (only

near-minimal). Does this distinction have the potential to bear a contrastive load? Further experimentation could test if American English speakers can use this distinction to recoverably contrast minimally paired nonce words.

Another extension of this study would be to analyze the acoustic character of glide-vowel distinctions in other languages, such as those documented by the many studies cited throughout this paper. This study's results are only intended to shine light on what representation may be most plausible (or at least rule any candidates out) for the distinction apparent in the American English phonological system under consideration. It is possible that languages previously argued on more phonological grounds to be best represented with the other approaches considered do actually cue it differently, with acoustic characterizations in line with those predicted by the respective representations. This approach of acoustic characterization is further applicable to the analysis of any distinction for which there is a diverse suite of potential acoustic cues. And, as employed here, that acoustic characterization may be useful in comparing the acoustic predictions generated by competing phonological representations of such a distinction and therefore speaking between them. Further such analysis will contribute to the ongoing broader question of how interwoven or disconnected phonological representation and phonetic realization can be (e.g., Pierrehumbert, 1990; Hayes et al., 2004; Smith, 2005).

Acknowledgements

I would like to thank Lisa Davidson, Maria Gouskova, and Frans Adriaans for their help and feedback at many stages of this research. Many additional thanks go to Susannah Levi, Suzy Ahn, Sean Martin, Becky Laturnus, members of the NYU Phonetics and Experimental Phonology Lab, the anonymous reviewers of this paper's earlier manuscript, and audiences at the 170th meeting of the Acoustical Society of America and the 2017 annual meeting of the Linguistic Society of America for valuable feedback and discussion.

Notes

[1]While the word-initial patterning of [jV] vs. [iV] appears to conflate with stress, one exception of [iV] hiatus where the following vowel, instead of the initial vowel, is stressed might be the name *Iago* [iάgo]. However, this name is not highly frequent in English and could be assigned something akin to a loan status, possibly allowing for phonological exceptionality (Itô and Mester, 1999; Smith, 2006; 2009).

[2]The pattern of [w] being dispreferred after labial consonants is not exceptionless. Some Spanish loanwords such as *Buena Vista* [bwɛnəvístə] maintain [w] after a labial consonant in their adaptations, though other loan adaptations do still exhibit this constraint, such as *Puerto Rico* [pɔɹɾəɹíko] or the variable adaptation of French *voila* as [wɑlά].

[3]Smoothing Spline ANOVA analysis was first used in linguistics by Davidson (2006) in the analysis of tongue shapes imaged by ultrasound to examine differences in coarticulation. De Decker and Nycz (2006) extended the use of this analytical tool to the study of vowel formants, finding this analysis of temporal formant contours to informatively reveal differences between vowel categories and dialectal category variants that single-point or timespan-averaged analyses might otherwise miss. Further work has employed this method in the analysis of vowels and diphthongs (Koops, 2010; Chanethom, 2011).

References

Aguilar, L. (1999). Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication*, 28(1):57–74. doi:10.1016/S0167-6393(99)00003-5

Ainsworth, W. (1974). The influence of precursive sequences on the perception of synthesized vowels. *Language and Speech*, 17(2):103–109. doi:10.1177/002383097401700201

Boersma, P. and Weenink, D. (2015). Praat: Doing phonetics by computer [computer program], version 5.3.77. URL http://www.fon.hum.uva.nl/praat/

Bright, W. (1957). *The Karok Language*. Berkeley and Los Angeles: University of California Press.

Browman, C. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J. and Beckman, M., editors, *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, pages 341–376. Cambridge University Press. doi:10.1017/CBO9780511627736.019

Catford, J. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Chanethom, V. (2011). Dynamic differences in the production of diphthongs by French-English bilingual children. *The Journal of the Acoustical Society of America*, 130(4):2522–2522. doi:10.1121/1.3655063

Chitoran, I. (2002). A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association*, 32(2):203–222. doi:10.1017/S0025100302001044

Clements, G. N. and Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Number 9. MIT Press, Cambridge, MA.

Crystal, T. and House, A. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, 88(1):101–112. doi:10.1121/1.399955

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, 120(1):407–415. doi:10.1121/1.2205133

Davidson, L. and Erker, D. (2014). Hiatus resolution in American English: The case against glide insertion. *Language*, 90(2):482–514. doi:10.1353/lan.2014.0028

Davis, S. and Hammond, M. (1995). On the status of onglides in American English. *Phonology*, 12:159–182. doi:10.1017/S0952675700002463

De Decker, P. and Nycz, J. (2006). A new way of analyzing vowels: Comparing formant contours using Smoothing Spline ANOVA. *Poster presented at New Ways of Analyzing Variation (NWAV) 35, 9-12 November, Columbus, Ohio*.

Deligiorgis, I. (1988). *Glides and syllables*. PhD thesis, University of Iowa.

Durand, J. (1987). On the phonological status of glides: The evidence from Malay. In Anderson, J. and Durand, J., editors, *Explorations in Dependency Phonology*, pages 79–107. Dordrecht, Holland: Foris Publications.

Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels/wjrl/in American English. *The Journal of the Acoustical Society of America*, 92(2):736–757. doi:10.1121/1.403998

Fowler, C., Brown, J., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3):396–413. doi:10.1016/S0749-596X(03)00072-X

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, 20(2):269–337. doi:10.1023/A:1014942312445

Gay, T. (1968). Effect of speaking rate on diphthong formant movements. *The Journal of the Acoustical Society of America*, 44(6):1570–1573. doi:10.1121/1.1911298

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America*, 63(1):223–230. doi:10.1121/1.381717

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, 38(1-3):148–158. doi:10.1159/000260020

Gentilucci, M. and Bernardis, P. (2007). Imitation during phoneme production. *Neuropsychologia*, 45(3):608–615. doi:10.1016/j.neuropsychologia.2006.04.004

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association*, 32(2):113–121. doi:10.1017/S0025100302001007

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2):251–279. doi:10.1037/0033-295X.105.2.251

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Gouskova, M. (2004). Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology*, 21(02):201–250. doi:10.1017/S095267570400020X

Grosjean, F. and Lane, H. (1976). How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4):538–543. doi:10.1037/0096-1523.2.4.538

Gu, C. (2002). *Smoothing Spline ANOVA Models*. New York: Springer. doi:10.1007/978-1-4614-5369-7

Halle, M., Vaux, B., and Wolfe, A. (2000). On feature spreading and the representation of place of articulation. *Linguistic Inquiry*, 31(3):387–444. doi:10.1162/002438900554398

Harris, J. and Kaisse, E. (1999). Palatal vowels, glides, and obstruents in Argentinean Spanish. *Phonology*, 16:117–190. doi:10.1017/S0952675799003735

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2):253–306.

Hayes, B., Kirchner, R., and Steriade, D., editors (2004). *Phonetically based phonology*. Cambridge University Press.

Hirata, Y. (1990). Perception of geminated stops in Japanese word and sentence levels. *The bulletin of the Phonetic Society of Japan*, 194:23–28.

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, 32(4):565–589. doi:10.1016/j.wocn.2004.02.004

Hyman, L. (1985). *A theory of phonological weight*. Dordrecht, Holland: Foris Publications.

Itô, J. and Mester, A. (1999). The phonological lexicon. In Tsujimura, N., editor, *Handbook of Japanese Linguistics*, pages 62–100. Oxford: Blackwell.

Jensen, J. T. (1993). *English Phonology*, chapter Segmental Phonology, pages 25–45. Amsterdam: John Benjamins. doi:10.1075/cilt.99

Kaye, J. and Lowenstamm, J. (1984). De la syllabicité. In Dell, F., Hirst, D., and Vergnaud, J.-R., editors, *Forme sonore du langage*, pages 123–159. Paris: Hermann.

Keating, P. (1988). Palatals as complex segments: X-ray evidence. *UCLA working papers in phonetics*, 69:77–91.

Klatt, D. (1973). Interaction between two factors that influence vowel duration. *The journal of the Acoustical Society of America*, 54(4):1102–1104. doi:10.1121/1.1914322

Klatt, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3):129–140.

Koops, C. (2010). /u/-fronting is not monolithic: Two types of fronted /u/ in Houston Anglos. *University of Pennsylvania Working Papers in Linguistics*, 16(2):14.

Kozchevnikov, V. and Chistovich, L. (1965). *Speech: Articulation and perception* [translated]. Washington, DC: Joint Publications Research Service.

Levi, S. V. (2004). *The representation of underlying glides: A cross-linguistic study*. PhD thesis, University of Washington.

Levi, S. V. (2008). Phonemic vs. derived glides. *Lingua*, 118(12):1956–1978. doi:10.1016/j.lingua.2007.10.003

Levin, J. (1985). *A metrical theory of syllabicity*. PhD thesis, Massachusetts Institute of Technology.

Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of experimental psychology*, 52(2):127–137. doi:10.1037/h0041240

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35(11):1773–1781. doi:10.1121/1.1918816

Lindblom, B. (1968). Temporal organization of syllable production. *Speech transmission laboratory quarterly progress and status report*, 2(3).

Maddieson, I. (2008). Glides and gemination. *Lingua*, 118(12):1926–1936. doi:10.1016/j.lingua.2007.10.005

Maddieson, I. and Emmorey, K. (1985). Relationship between semivowels and vowels: Cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica*, 42(4):163–174. doi:10.1159/000261748

Namy, L., Nygaard, L., and Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21(4):422–432. doi:10.1177/026192702237958

Nevins, A. and Chitoran, I. (2008). Phonological representations and the variable patterning of glides. *Lingua*, 118:1979–1997. doi:10.1016/j.lingua.2007.10.006

Ohala, J. J. (1978). Southern Bantu vs. the world: The case of palatalization of labials. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 4:370–386. doi:10.3765/bls.v4i0.2218

Paradis, C. (1992). *Lexical phonology and morphology: The nominal classes in Fula*. New York: Garland Publishing Inc.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of phonetics*, 18(3):375–394.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2017). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-131.

RCoreTeam (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ren, H. (1986). On the acoustic structure of diphthongal syllables. In *UCLA Working Papers in Phonetics*, volume 65. UCLA.

Roca, I. (1997). There are no 'glides', at least in Spanish: An optimality account. *Probus*, 9:233–265. doi:10.1515/prbs.1997.9.3.233

Rosenthall, S. (1994). *Vowel/glide alternation in a theory of constraint interaction*. PhD thesis, University of Massachusetts-Amherst.

Smith, J. (2003). Onset sonority constraints and subsyllabic structure. In Rennison, J. R., Pöchtrager, M. A., and Neubarth, F., editors, *Phonologica 2002*. Berlin: de Gruyter.

Smith, J. (2005). Phonological constraints are not directly phonetic. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 41, pages 457–471. Chicago Linguistic Society.

Smith, J. L. (2006). Loan phonology is not all perception: Evidence from Japanese loan doublets. In Vance, T. J. and Jones, K. A., editors, *Japanese/Korean Linguistics 14*, pages 63–74.

Smith, J. L. (2009). Source similarity in loanword adaptation: Correspondence Theory and the posited source-language representation. In Parker, S., editor, *Phonological Argumentation: Essays on Evidence and Motivation*, pages 155–177. London: Equinox.

Steriade, D. (1984). Glides and vowels in Romanian. In Brugman, C. and Macaulay, M., editors, *Proceedings of the 10th Annual Meeting of the Berkeley Linguistics Society*, pages 47–64. doi:10.3765/bls.v10i0.1935

Steriade, D. (1988). Review of Clements & Keyser (1983). *Language*, 64:118–129. doi:10.2307/414790

Stone, M. (2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, 19:455–502. doi:10.1080/02699200500113558

Straka, G. (1964). A propos de la question des semi-voyelles. *STUF - Language Typology and Universals*, 17:301–323. doi:10.1524/stuf.1964.17.16.301

Townsend, C. and Janda, L. (1996). *Common and comparative Slavic: Phonology and inflection*. Columbus, Ohio: Slavica.

Turner, G. S., Tjaden, K., and Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, 38(5):1001–1013. doi:10.1044/jshr.3805.1001