DE GRUYTER MOUTON
DOI 10.1515/lp-2014-0010 — Laboratory Phonology 2014; 5(2): 271–288

Aude Noiray*, Khalil Iskarous and D. H. Whalen

Variability in English vowels is comparable in articulation and acoustics

Abstract: The nature of the links between speech production and perception has been the subject of longstanding debate. The present study investigated the articulatory parameter of tongue height and the acoustic F1–F0 difference for the phonological distinction of vowel height in American English front vowels. Multiple repetitions of /i, ɪ, e, ɛ, æ/ in [(h)Vd] sequences were recorded in seven adult speakers. Articulatory (ultrasound) and acoustic data were collected simultaneously to provide a direct comparison of variability in vowel production in both domains. Results showed idiosyncratic patterns of articulation for contrasting the three front vowel pairs /i-ɪ/, /e-ɛ/, and /ɛ-æ/ across subjects, with the degree of variability in vowel articulation comparable to that observed in the acoustics for all seven participants. However, contrary to what was expected, some speakers showed reversals for tongue height for /ɪ/-/e/ that were also reflected in acoustics, with F1 higher for /ɪ/ than for /e/. The data suggest the phonological distinction of height is conveyed via speaker-specific articulatory-acoustic patterns that do not strictly match features descriptions. However, the acoustic signal is faithful to the articulatory configuration that generated it, carrying the crucial information for perceptual contrast.

***Corresponding author: Aude Noiray:** Haskins Laboratories. Also at Center for Excellence Cognitive Science, Potsdam University. E-mail: noiray@haskins.yale.edu
Khalil Iskarous: Haskins Laboratories. Also at Linguistic Department, University of Southern California. E-mail: kiskarou@usc.edu
D. H. Whalen: Haskins Laboratories. Also at Program in Speech-Language-Hearing Sciences, the Graduate Center, City University of New York. E-mail: whalen@haskins.yale.edu

1 Introduction

Finding the source(s) of distinctness in vowels has remained an object of vigorous debate for feature theory. For a long period of time, a likely reason for this controversy was the lack of sufficiently sophisticated instrumental methods supporting the direct observation of the articulatory mechanisms underlying acoustic contrasts in vowel production. Although the development of various articulatory data collection techniques (e.g., x-ray methods, electromagnetic articulometry

[EMA], dynamic MRI, and ultrasound imaging) has improved the experimental situation, finding reliable correspondences between phonological and phonetic descriptions of vowels has remained problematic within and across languages. However, such correspondences are essential for attesting the relevance of phonological features and for advancing our understanding of language systems.

In Trubetzkoy's (1969 [1939]) early work on phonological description and in later formulations (Jackobson et al. 1952; Chomsky and Halle 1968), phonemes are defined as *oppositions* of *distinctive features*. In the specific case of *height*, the features [+high] and [+low] allow for the opposition between [i] and [e] or [a]. Regarding the phonetic realization of height contrast, various studies initially proposed a one-to-one relation between articulation and phonological description. Examples can be found in the IPA system or the cardinal vowel system of Daniel Jones, in which an articulatory basis for the dimension of tongue height was presumed before physical measurements of the tongue's position that could assess the height ordering of vowels were available.

As observational methods have improved, a number of studies have examined the highest point of the tongue in the oral cavity to distinguish vowels by height (e.g., Jones 1966 [1917]; Abercrombie 1967; Fischer-Jørgensen 1985; Noiray et al. 2008). However, such methods have also led to the discovery of vowel reversals or 'flips' in several speakers (e.g., Russell's [1928] x-ray studies; Ladefoged [1962]; Ladefoged et al.'s [1972] cinefluorographic investigations; Wood's [1975, 1979] studies in various languages; Johnson et al. [1993], for which the physical highest point on the tongue did not corroborate the theoretically predicted height from phonological representations). Therefore, the finding in articulation of speaker-specific idiosyncrasies does not support a one-to-one correspondence between the conventional phonological representation of vowel height and the physical vertical position of the tongue in the oral cavity. The maintenance of contrast is still assumed, however, at least to the extent that these researchers judged the tokens of the vowels to have been acceptable exemplars of those categories.

Another possible analysis for the contrasts, then, would be a one-to-one relation between acoustic realizations and phonological representations of vowels. Such an account has been proposed for several languages: American English (Ladefoged et al. 1972; Nearey 1978; Stevens and Blumstein 1978; Lindblom 1986; Johnson et al. 1993; Perkell et al. 2000); Ningbo Chinese (Hu 2005); and French (Ménard et al. 2008). Certainly, generalizations have often been made about the correlation of vowel features and acoustics. Variation in height is associated with F1 (Joos 1948; Lindblom and Sundberg 1971) and fronting with change in F2 or as a difference between F2 and F1 (Fant 1960; Ladefoged 1975). On such accounts, the acoustic pattern of formants is therefore the target for each vowel, and the

exact means by which it is achieved is irrelevant (e.g., Ladefoged et al. 1972; Stevens and Blumstein 1978; Lindblom 1986; Perkell et al. 2000; Johnson 2012: 144).

Although there is often disagreement between the articulatory and acoustic approaches, inter-speaker and/or intra-speaker variability in both the acoustic and the articulatory specification of vowels has been consistently reported, even for nearly steady-state vowel productions (Russell 1928; Ladefoged et al. 1972; Wood 1975, 1982; Fischer-Jørgensen 1985; Johnson et al. 1993). Contextual, co-articulatory effects are found, even across intervening segments (e.g., Alfonso and Baer 1982; Cole et al. 2010). The sources of inter-speaker variability may be multiple: dialectal or regional (Clopper et al. 2005), anatomical (Stevens and House 1955; Perkell et al. 1997; Brunner et al. 2009), and/or related to idiosyncratic strategies. Intra-speaker variability has typically been attributed to peripheral factors (e.g., coarticulatory effects due to contextual variation, changing phoneme realization compared to how they would be produced sequentially; MacNeilage 1970), but there could conceivably be central origins as well (i.e., in intentional motor commands). Indeed, variability has been claimed to be a useful source of information for infants learning a language (Rost and McMurray 2010).

Finally, for both consonants and vowels, there is also a large literature on acoustic variability (e.g., Peterson and Barney 1952; Perkell and Klatt 1986; Hillenbrand et al. 1995). The extent of acoustic variability has generated its own literature on acoustic normalization to try to account for perceptual constancy (Disner 1980; Adank et al. 2004; Clopper 2009; Flynn 2011). But even when vowel acoustics are mapped onto auditory scales approximating listener perception, there can be extensive intra-speaker and inter-speaker variability in vowel acoustics, as shown by the data of a recent study (Ménard et al. 2008).

The present paper addresses the debate between the articulatory and acoustic approaches on vowel production with two questions:

1. Is there more variability in the articulatory domain than in the acoustics such that the acoustics is what conveys vowels distinctiveness?
2. Do speaker-specific idiosyncrasies in the articulatory domain have consequences in the acoustic domain?

We addressed these questions using a methodology that allows us to simultaneously record speech articulation and acoustics with high accuracy, and an experimental design which allows us to examine various aspects of variability in the two domains. More specifically, we examined two main articulatory and acoustic correlates of the dimension of height in the American English front vowels, i.e., tongue height and the F1–F0 difference. The data set also allows us to test the Ladefoged et al. (1972) observation that the highest point of the tongue for the

vowels /ɪ/ and /e/ is flipped by many speakers. This observation is used to argue that these *flips* make F1 and F2 (or some transform of them) better descriptors of vowel features than the highest point of the tongue, since the acoustic parameters would be more consistent in their ordering than the articulatory ones. However, Ladefoged et al. (1972) do not discuss whether the acoustic parameters for these subjects also flip. The data we present allow us to examine this issue. We used an environment, [hVd] sequences, which results in very little coarticulatory formant movements (Stevens and House 1963).

2 Method

2.1 Participants

Seven monolingual adult speakers of American English (three males, four females, aged 20–35) were recruited in Connecticut to participate in the experiment. Four of the participants were native to Connecticut, and the other three had lived in Connecticut for at least 6 years, starting at ages 4, 9, and 21 years. None reported any history of hearing deficits or cognitive or motor disorders. All were compensated for their participation in the study.

2.2 Stimuli

The target vowels /i, ɪ, e, ɛ, æ/ were embedded in [(h)Vd] real words (*heed*, *hid*, *head*, *aid*, and *had*) to obtain nearly steady-state vowels minimizing coarticulatory effects. They were presented as written prompts on a computer screen. Participants produced 15 repetitions of each target sequence. Sequences were displayed in randomized blocks using an in-house program. The randomization procedure aimed to avoid any bias on the acoustic and articulatory data collected due to possible habituation effects (e.g., avoiding subjects' predicting subsequent stimuli in the case of a limited list of stimuli). Stimuli occurred at approximately 3-second intervals. Disfluencies were excluded; these constituted less than 1% of the utterances. The protocol was short, so that fatigue was not a factor.

2.3 Experimental procedure

Prior to the recording, participants were familiarized with the list of stimuli to ensure they had no difficulty producing the target words. Participants were in-

structed to produce each word at their natural rate. A short break was made between each block.

Both the acoustic speech signal and articulatory data from the tongue were collected using a combined digital ultrasound system imaging the tongue on the midsagittal plane at a high sampling rate (127 Hz) with three-dimensional optical tracking of the head and ultrasound probe (Optotrak, Northern Digital), allowing for relatively unconstrained motion of the head and jaw during speech (Whalen et al. 2005).

In this study, we used an Aloka SSD-5500 ultrasound unit with an intercostal probe operating at 7.5 Mhz, an angle of 90 degrees, and 17 cm depth. Ultrasound imaging is non-invasive and provides real-time measurements of the tongue during various speech tasks. It is possible to see most of the tongue surface, from the upper root to near the tip, so long as the tongue surface remains visible. The highest point of the tongue is always captured, while in point-tracking techniques (e.g., EMA, x-ray microbeam where tracking points are added along the midline of the midsagittal tongue), the actual highest point could occur between measured points and thus be somewhat underestimated. In a comparison between EMA and ultrasound imaging accuracy for assessing tongue motion, Honda and Kaburagi (1993) found an average measurement error of 1.16 ±0.74 mm for "the distance between the magnetically measured position and the ultrasonic tongue contour". Given that the pixel size in ultrasound images is approximately 1 mm, this is essentially as accurate as the system can be. The effects of the correction for movement of the head and the probe can be seen in Figure 1.

In this study, head motion was tracked via a headgear incorporating six small infrared emitting diodes (IREDs). This procedure was designed to subsequently re-express tongue motion in head coordinates (Whalen et al. 2005). Five IREDs were glued on a plate coupled with the ultrasound probe to track tongue position relative to the head (Figure 2). Correction of the tongue edge position was subsequently conducted via pitch rotation, horizontal as well as vertical translation in the direction of the motion of the probe (see Ostry et al. 1996 for details).

The occlusal plane was determined for each participant by simultaneously recording the position of the IREDS on the headgear and of a Plexiglas triangle clenched between the teeth of the participants, on which three Optotrak IREDs were glued. The occlusal plane was used as a reference for subsequent rotation and correction of the tongue data and the hard palate structure (Westbury et al. 2002).

To image the hard palate structure, a 'swallowing' trial was collected. The speaker was instructed to press a water bolus toward the hard palate, which allowed the ultrasound signal to reach and reflect the hard palate. To avoid possible air pockets at the top of the bolus, the speaker then swallowed the bolus

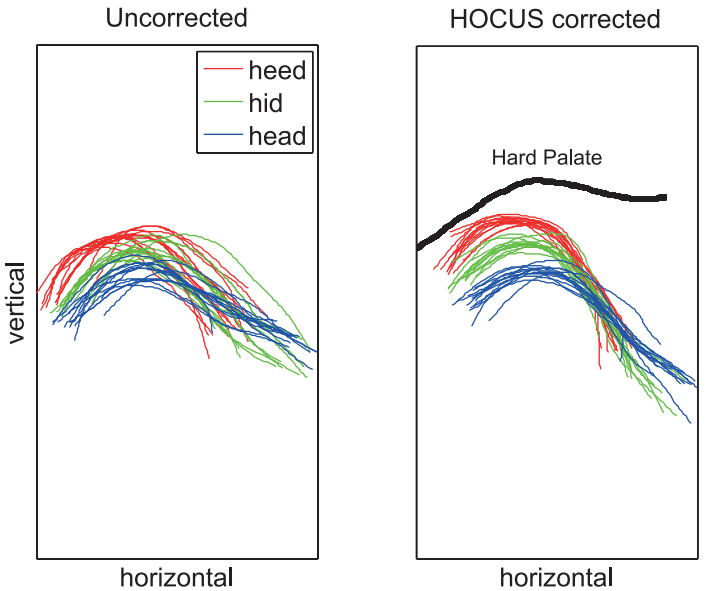

Fig. 1: Midsagittal views of uncorrected tongue edges in jaw frame (left panel) and corrected tongue edges in head frame (right panel) for an American English speaker after rotation and correction of the tongue data. The vowels /i, ɪ, ɛ/ are represented in gray scale. The anterior portion of the tongue is on the left, the posterior portion on the right. The x-axis represents the horizontal axis along the front-back dimension within the oral cavity (in mm). We aligned the horizontal axis to each speaker's occlusal plane. Also, the y-axis characterizes the position of the tongue on height dimension (in mm). The bold line above the tongue edges represents the midsagittal palate trace from the swallow trial (labeled Hard Palate).

so that the entire length of the palate would be imaged at some point during the trial. This procedure was conducted at the beginning of the session. The palate image was applied to the subsequent images of the tongue collected during speech and served as a reference point for the tongue position in the x-axis in the oral cavity.

Participants were seated comfortably in an adjustable chair. The ultrasound probe was held in a spring-loaded probe holder fixed to a weighted customized pedestal (Figure 2). These have been designed to adapt the set-up to participant morphology and to provide a comfortable environment. The probe holder permits the ultrasound probe to move smoothly along the vertical axis with the natural downward motion of the jaw rather than being fixed in position as in other systems. However, while motion in the vertical dimension is possible, the probe holder prevents the probe from moving along the lateral and horizontal axes. This allows us to obtain tongue contours on the midsagittal plane across repetitions.

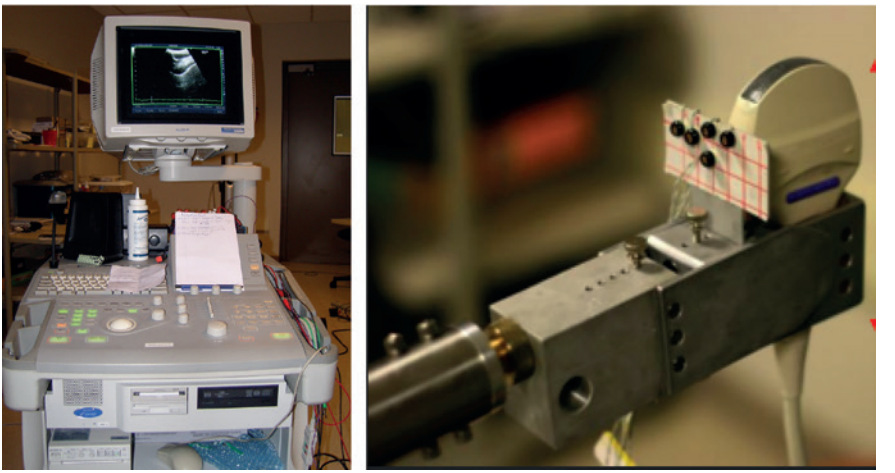

Fig. 2: Ultrasound unit (left panel); probe holder and probe (right panel) used for recording participants. The customized design includes five IREDs for tracking displacement of the probe along the vertical dimension.

The acoustic speech signal was recorded using a head-mounted microphone (Audio-Technica Model ATM 75) at 22.05 kHz. The acoustic and spatial data were simultaneously recorded on the Optotrak unit. The synchronization of the two systems was performed using a transistor–transistor logic (TTL) signal (to the Optotrak system) and a simultaneous signal to the ultrasound machine. This consisted of a series of trigger pulses simulating electrocardiogram (ECG) signals, which is the only external synchronization available with this device.

2.4 Post processing

For each sequence, ultrasound images were extracted and enhanced using Matlab (The MathWorks). Tongue contours were extracted using a semi-automatic detection system (EdgeTrak; Li et al. 2003), which resulted in a curve of 100 points.

Recalculating the position of the tongue edges from a probe-based to head-based coordinate space was done via optical tracking of 3D IREDS. For each sequence, optical tracking via Optotrak allowed for the localization in space of the five IREDs on the probe and the six IREDS on the headgear to obtain probe and head position (Whalen et al. 2005). To determine the rigid body coordinates (translations and rotations) of the head and probe, a set of MATLAB procedures

previously developed for jaw motion detection (Ostry et al. 1996) was used. A two-step optimization procedure was first used to correct for head motion relative to the Optotrak camera and specify the motion of the probe in a head-centered coordinate system for that frame.

For each target ultrasound frame, the rigid body reconstruction and correction procedures determine six numbers specifying the position and orientation of the probe for that frame. Three of these specify the vertical, lateral, and horizontal position, and the other three specify the pitch, roll, and yaw. Such correction allows for compensation for head motion and change of orientation of the probe. For each frame of interest, tongue contours are then corrected to be relative to the palate in a coordinate system aligned with the occlusal plane.

2.5 Analyses

For each participant and target [(h)Vd] sequence, utterances were first judged to be the correct target vowel by the experimenters and discarded if not. Speakers correctly produced all stimuli. For each token, several measurements were made to provide an acoustic and articulatory characterization of front vowels. On the acoustic speech signal, we conducted an LPC analysis for each vowel (30 ms centered at the midpoint of the vowel, 12 coefficients) as well as an automatic peak-picking algorithm on the spectrum to obtain F0 and F1. A Hamming window and preemphasis were applied before formant extraction. The F1–F0 difference was considered an indicator of vowel height (Traunmüller 1981; Syrdal and Gopal 1986); this is presumably accurate only when the items are in similar prosodic contexts, as they were here. Note that for the token *aid,* formant frequencies were measured during the /e/ portion of the diphthong, before any possible /i/ off-glide (though the majority of our utterances did not have noticeable off-glides). This point was approximately 30% through the vocalic segment. The main reason for using F1–F0, rather than just F1, is that the former measure is an auditory measure, which has been argued to be more perceptually useful to listeners as an indicator of vowel height (Syrdal and Gopal 1986). One possible problem with this measure is that if the different vowels were produced in different prosodic conditions (due to list intonation, for instance), then F1–F0 would reflect prosodic factors as well as the height-related factors of vowel height and intrinsic pitch. The vowels in this experiment were randomized to avoid a prosodic confound. To test for the possibility of such a confound, we conducted a mixed-effects test with the dependent variable F0 and independent fixed effect Vowel, and Subject as a random effect. The baseline for the contrast was the vowel /æ/. As expected, /i/ was on average 22 Hz higher than /æ/, with a standard error of 3 Hz, and /ɪ/ was

11 Hz higher, also with a standard error of 3 Hz. These are typologically expected intrinsic pitch effects related to vowel height (Whalen and Levitt 1995); therefore, a prosodic confound does not seem to be present in the data.

On the ultrasound image of the tongue shape corresponding to the acoustic midpoint of the vocalic segment, the highest point of the tongue dorsum was measured in the coordinative system adjusted by the optical tracking as an estimate of tongue height. The location of this point in the anterior-posterior dimension was taken as an indicator of tongue advancement.

These measures were made for each repetition of the vowel (/i, ɪ, e, ɛ, æ/).

3 Results

3.1 Articulatory and acoustic distinctness

The notion that vowels are more acoustically separable than they are articulatorily for the phonological feature of height in American English is examined here by quantifying distinctiveness in the two domains.

To do so, we used Cohen's *d* measure that can be used in addition to tests of significance to measure the amount of difference (i.e., size effect) between two distributions that are measured on different scales (Cohen 1992); in our case, the two groups of data are height values measured in millimeters and formant frequencies measured in Hertz.

Specifically, Cohen's *d* measures the distance between means of two distributions divided by the pooled variability of the two distributions: $d = \frac{\bar{x}_1 - \bar{x}_2}{s}$. In this study, we used Cohen's *d* as a way of normalizing differences across participants with different-sized vocal tracts. Cohen's *d* was calculated for F1–F0 and for the height of the highest point of the tongue (VH) for each contiguous pair of vowels and each subject.

A Cohen's *d* value of .8 indicates relatively high distinctiveness (or low dispersion) in standard deviations, while a value of .2 indicates lesser distinctiveness. A linear mixed-effects test of a significant difference between acoustically and articulatorily measured Cohen's *d* was performed. The dependent variable was Cohen's *d* magnitude, the independent variable was AcArt (Levels: Articulatory, Acoustic), and the random effect was Subject (random intercept). Acoustic Cohen's *d* was estimated to be .62, which was higher than the articulatory Cohen's *d*, but the standard error was .82, which is quite large, with $t(6) = .82$ ($p > .05$). Therefore, there was no significant difference between the two Cohen's *d* (even with this non-conservative estimate of degrees of freedom).

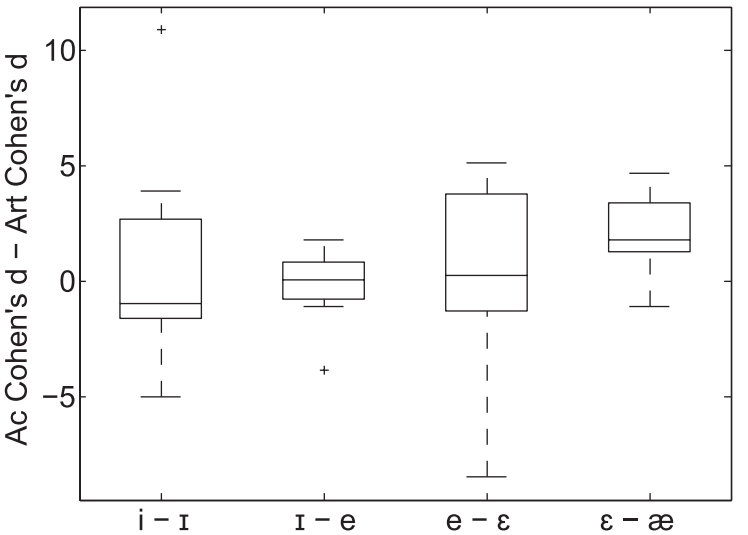

Fig. 3: Results for Cohen's *d* measures calculated across subjects for acoustic vs. articulatory distinctness for each vowel pair.

One possibility is that some particular vowel pairs show greater acoustic distinctness (estimated as the difference in Cohen's *d* magnitudes) than articulatory distinctness, while others do not. Of the 28 comparisons (7 subjects × 4 vowel pairs), there were 13 cases in which acoustic distinctness was greater than articulatory distinctness, 10 cases in which articulatory distinctness was greater than acoustic, and 5 cases in which the two were less than .5 standard deviations from each other. The boxplots in Figure 3 show the distribution across subjects of magnitude of the acoustic Cohen's *d* minus the magnitude of the articulatory Cohen's *d*, as a function of vowel category. Only the /ɛ/–/æ/ category shows consistently greater acoustic distinctness than articulatory distinctness, with 6 of the 7 subjects showing the pattern. A *t*-test was performed for each of the categories to test if the mean was different from 0. Only the /ɛ/–/æ/ pair shows a significant result with $t(6) = 2.71$, $p < .05$.

3.2 Vowel flips

One of the first explicit arguments for considering F1 and F2, or some nonlinear function of them, rather than the highest point of the tongue to be the primary distinguishing features for vowels was the Ladefoged et al. (1972) finding that for many subjects, the highest point of the tongue is lower for /ɪ/ than for /e/,

counter to expectations from phonology. However, as far as we know, there has never been a thorough examination of whether subjects that flip their articulatory heights for /ɪ/ and /e/ also flip the F1 of each vowel. If the latter happens, height flipping cannot be used as an argument for the inadequacy of articulatory features. To examine whether subjects who flip /ɪ/ and /e/ articulatorily also flip them acoustically, we calculated Cohen's *d* based on F1–F0, an auditory indicator of vowel height (when prosodic environments are comparable), and on the height of the highest point of the tongue. The acoustic Cohen's *d*, therefore, calculates the distance between the distributions of F1–F0 for the two vowels, whereas the articulatory Cohen's *d* measures the distance between the distributions of vowel heights for those same vowels. Thus the two measures are independent and allow us to see whether flips in one domain correspond to flips in the other. In the units used here, a positive Cohen's *d* indicates that /ɪ/ is higher than /e/. The upper panel of Figure 4 presents Cohen's *d* for the participants that do not flip, and the lower panel shows those for the participants that flip. The horizontal dashed line indicates a Cohen's *d* of .8, which is considered to indicate 'large' effects (Cohen 1992). It can be seen that participants that flip articulatorily also flip acoustically. That is, when the highest point of the tongue for /ɪ/ is lower than that for /e/, the acoustic indicator of Height (F1–F0) is also higher for /ɪ/ than for /e/. However, it can be seen from Figure 4 that the articulatory and acoustic inter-distributions are only qualitatively similar, not quantitatively equal. Our statements are not about the fine details of the articulatory-acoustic correspondence, which are influenced by aspects of the articulatory-acoustic mapping that are specific to each subject and that may not be accounted for by a linear source-filter theory. For instance, Subject 5's articulatory difference is much higher than the acoustic difference, whereas Subject 2's acoustic difference is larger than the articulatory. And while for Subject 4 there's a considerable articulatory difference, there is a marginal acoustic difference in the reverse direction. In this paper we offer no explanation for these fine individual differences in the articulatory-acoustic maps, but we believe that this needs to be further investigated.

4 Discussion

This study investigated the relationship between articulatory and acoustic variability in vowel production. Our two main research questions were:

1. Is height distinctiveness better conveyed in the acoustic than in the articulatory domain?
2. Do speaker-specific idiosyncrasies in the articulatory domain affect the acoustic domain?

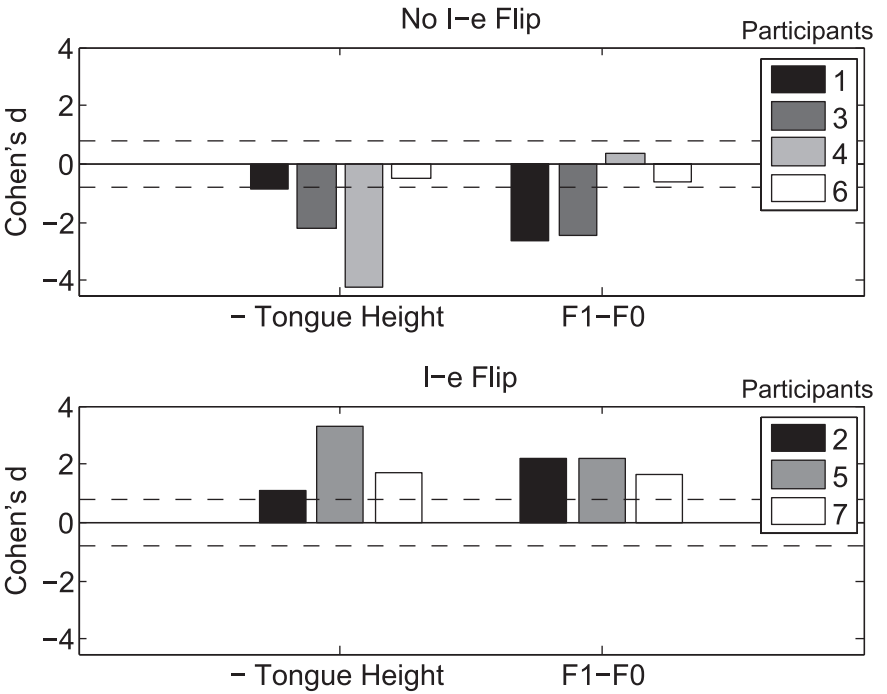

Fig. 4: Results for Cohen's *d* measures calculated across subjects (grayscale) for tongue height and F1–F0 distance for the vowel pairs /ɪ-e/. The upper panel shows results for participants who do not exhibit a flip between the two vowels, while the bottom panel presents results for those showing flips in height between the two vowels. The speakers with 'flips' had all spent their first 4–21 years in another state (Wisconsin, Indiana, or Pennsylvania). Those without flips had been born in Connecticut and lived there consistently.

Multiple repetitions of /i, ɪ, e, ɛ, æ/ in [(h)Vd] sequences were recorded in seven adult speakers of American English. Tongue height as well as the acoustic speech signal were simultaneously collected with optical-ultrasound systems and audio recording equipment.

4.1 Consistency between articulatory and acoustic variability

This study showed that the seven participants exhibited idiosyncratic patterns of variability in the articulatory domain as had been found previously (e.g., Ladefoged et al. 1972; Johnson et al. 1993), but, in a result not reported before, the variability in articulation was reflected in the acoustics. This is an important

result that contradicts studies that have only investigated articulatory variability while speculating on the acoustic variance-reduction. For example, Johnson et al. (1993: 713) concluded that "individual differences such as these (as well as vowel differences reported above) may be interpreted as indirect evidence that the acoustic product of speaking is the crucial determinant of the organization of speech articulation". In the present study, the comparison of the articulatory and acoustic domains provides direct evidence that individual variability in speech articulation has consequences in acoustics despite previous claims that speakers exhibit stable acoustic targets with little impact from articulatory variability during vowel production (e.g., Ladefoged et al. 1972; Johnson et al. 1993; Perkell et al. 1993; Savariaux et al. 1995; Brunner 2008). Further, these data show that in steady-state production of vowels, acoustic and articulatory distinctiveness do not differ significantly, contrary to previous claims. This result supports a conclusion of Maeda (1991: 328, Table 1) who found an acoustic reduction of variability in the production of coarticulated vowels by two French speakers but comparable variability in both domains once coarticulation effects were subtracted.

4.2 Vowel reversals

Previous articulatory studies, (e.g., Ladefoged et al. 1972) have suggested from vowel tongue height reversals or 'flips' observed in articulation that acoustic parameters (F1 and F2) are better descriptors of vowel height. However, these studies did not provide any actual support for this claim as they did not check for the presence of these vowel reversals in the acoustics. In the present investigation, we found similar articulatory patterns for /ɪ/ versus /e/ with three speakers producing the front vowel /e/ with a higher tongue position than for /ɪ/. Note that this value reflects both the tongue and the jaw height taken together, as in previous studies (e.g., Ladefoged et al. 1972). However, our results indicate that only the study of the articulation and the acoustics together can locate the discrepancies found between phonological representations of vowels and their phonetic realizations. The data indeed show that the articulatory reversals are reflected as reversals in the acoustics. This finding reinforces the hypothesis that there is a lawful relationship between acoustics and articulation in which the individual variability in articulation is directly retrieved in the acoustic signal and presumably available to listeners. Other aspects of this vowel pair (formant movement, duration) maintain the distinction (see Table 1). The phonological contrast of height may be better expressed in acoustics or articulation depending on the opposed vowel pair. Our results indicate that for /ɪ/ versus /e/, both the acoustic and

articulatory parameters concur on the height distinction. These data provide a compelling example that both sources of information are fundamental to phonological representation as they convey the distinctive properties of phonemes. Also, these results suggest that phonological distinction is achieved via idiosyncratic adjustments to produce perceptually relevant contrasts between vowels. Another interpretation could be that the individual patterns result from dialectal differences, or else that they evidence a disruptive response to changing dialect areas (either early or late in the speakers' lives). In our case, the four who did not flip were from Connecticut, while those who did were from Wisconsin, Indiana, or Pennsylvania. However, more speakers – with control for their sociolinguistic backgrounds – are needed to test such a hypothesis.

The next step is to determine what it is (off-glide, amplitude, duration) that maintains this distinction for these speakers. The experimenters heard the two vowels as intended, so this distinction was evidently maintained in perception. There was not any noticeable off-glide (i.e., diphthongization in /e/) in any of the three speakers showing the height reversals. The lack of diphthongization has been noted in the speech of younger generations in three dialect areas across the United States (Jacewicz et al. 2011), so this result could be due to a relatively recent change in the dialects of our speakers as well.

Both the speakers with flips and those without produced vowels that were recognizable as the intended vowel, so the reversed F1 must be overridden by other factors. All speakers had longer durations for /e/ than for /ɪ/, with a ratio of about 1.4 to 1 (see Table 1). In addition, there were differences in the formant trajectories, with F1 and F2 of /ɪ/ converging slightly over time, while /e/ was either flat or slightly diverging. Thus distinctiveness was maintained, even though a single point of measurement would lead us to expect confusion.

Table 1: Acoustic characteristics of /ɪ/ and /e/. Those speakers with reversed tongue height ('flippers' from the articulatory domain) are marked with shading and 'fl'.

Subject	Duration (s)		F0 (Hz)		F1 (Hz)		F2 (Hz)		F1–F0 (Hz)		dur ratio
	hid	aid	hid	aid	hid	aid	hid	aid	hid	aid	aid/hid
P001	0.1803	0.2433	219	203	478	498	2379	2497	259	295	1.35
P002 fl	0.2372	0.3378	153	137	460	394	2054	2349	307	257	1.42
P003	0.1957	0.2535	205	180	478	381	2264	2360	273	201	1.30
P004	0.1932	0.2813	122	115	387	345	2157	2320	266	229	1.46
P005 fl	0.1955	0.2575	235	206	522	456	2566	2889	287	250	1.32
P006	0.1750	0.2813	108	95	411	412	1950	2082	302	317	1.61
P007 fl	0.2017	0.2660	225	195	511	396	2144	2429	286	201	1.32

5 Conclusion

This study examined articulatory and acoustic correlates for height distinction in American English front vowels. The results suggest that speakers use both acoustics and articulation to express contrasts in vowel height. We found that front vowels were produced within a continuum of variability but that this variability was structured in each speaker to signal differences in vowel category (e.g., vowel reversals between /ɪ/ and /e/).

Results also indicate that the acoustic signal that carries the majority of the information is directly structured by articulation, and thus that most of articulation seems to be present in the details of the acoustics and should be available to listeners. In particular, the 'flippers' were no less comprehensible than the non-flippers, even though their F1/F2 values were flipped as well as the tongue height (contra the assumption of Johnson et al. 1993). This is no doubt partly due to duration differences (see Table 1), but may be more due to the movement of the formants throughout the vocalic segment. The vowels, even in this acoustically stable consonant context, had formant movements throughout, whether the vowel was nominally a diphthong (/e/) or not (/ɪ/). Even in this relatively stable environment, vowel formant dynamics appear to be perceptually salient (e.g., Strange et al. 1983). Iskarous, Nam, and Whalen (2010) found listeners to be sensitive to the movement patterns of formants (presumably reflecting the underlying kinematics) even when the vowel categories were always correctly identified. Such results argue for a theory of vowel perception that is sensitive to a great deal of the structure throughout the syllable. They are incompatible with any theory that depends on simple acoustic measures at a single time point, no matter where that point is selected.

Future work should examine to what extent vowel perception is sensitive to individual production variability in the articulatory and acoustic domains and also whether the information provided in articulation can be fully recovered by listeners via the acoustics. The exact nature of this information is clear neither from the present results nor from previous work. Vowel formant spaces capture a great deal of the linguistically significant acoustic structure of the world's languages, as we have known for some time. We do not currently have models capable of providing an explanation of the remaining, currently undescribed, structure. Our assumption is that listeners are recovering the underlying dynamics signaled by the acoustics, but current tools have not proven adequate to model this process. Nonlinear systems are intrinsically difficult to model (e.g., Ljung 2010), and the possible positive role of variability (e.g., Riley and Turvey 2002) makes investigation of large data sets necessary. Given the current paucity of articulatory data (especially compared with the volume of acoustic data), it is not

surprising that our theorizing is at an early stage. The increasing ease of collecting ultrasound, magnetic resonance imaging, electromagnetic articulography, and other physiological measures holds the promise of our eventually matching our theories to the capabilities and processes of human listeners.

References

Abercrombie, David. 1967. *Elements of general phonetics.* Edinburgh: Edinburgh University Press.

Adank, Patti, Roel Smits, & Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116. 3099–3107.

Alfonso, Peter J., & Thomas Baer. 1982. Dynamics of vowel articulation. *Language and Speech* 25. 151–173.

Brunner, Jana. 2008. Acoustic compensation and articulo-motor reorganisation in perturbed speech, Berlin: Humboldt-Universität zu Berlin dissertation.

Brunner, Jana, Susanne Fuchs, & Pascal Perrier. 2009. On the relationship between palate shape and articulatory behavior. *Journal of the Acoustical Society of America* 125. 3936–3949.

Chomsky, Noam, & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Clopper, Cynthia G. 2009. Computational methods for normalizing acoustic vowel data for talker differences. *Language and Linguistics Compass* 3. 1430–1442.

Clopper, Cynthia G., David B. Pisoni, & Kenneth de Jong. 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America* 118. 1661–1676.

Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112. 155–159.

Cole, Jennifer, Gary Linebaugh, Cheyenne Munson, & Bob McMurray. 2010. Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics* 38. 167–184.

Disner, Sandra Ferrari. 1980. Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America* 67. 253–261.

Fant, Gunnar. 1960. *Acoustic theory of speech production.* The Hague: Mouton.

Fischer-Jørgensen, Eli. 1985. Some vowel features, their articulatory correlates, and their explanatory power in phonology. In Victoria Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged,* 79–89. New York: Academic Press.

Flynn, Nicholas. 2011. *Comparing vowel formant normalisation procedures* (York Papers in Linguistics Series 2), 1–28.

Hillenbrand, James M., Laura A. Getty, Michael J. Clark, & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97. 3099–3111.

Honda, Masaaki, & Tokihiko Kaburagi. 1993. Comparison of electromagnetic and ultrasonic techniques for monitoring tongue motion. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation, München FIPKM* 31. 121–136.

Hu, F. 2005. A phonetic study of the vowels of Ningbo Chinese. Hong Kong: City University of Hong Kong Ph.D. dissertation.

Iskarous, Khalil, Hosung Nam, & D. H. Whalen. 2010. Perception of articulatory dynamics from acoustic signatures. *Journal of the Acoustical Society of America* 127. 3717–3728.
Jacewicz, Ewa, Robert A. Fox, & Joseph Salmons. 2011. Vowel change across three age groups of speakers in three regional varieties of American English. *Journal of Phonetics* 39. 683–693.
Jakobson, Roman, Gunnar Fant, & Morris Halle. 1952. *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.
Johnson, Keith. 2012. *Acoustic and auditory phonetics* (3rd ed). Malden, MA: Wiley-Blackwell.
Johnson, Keith, Peter Ladefoged, & Mona Lindau. 1993. Individual differences in vowel production. *Journal of the Acoustical Society of America* 94. 701–714.
Jones, Daniel. 1966 [1917]. *English Pronouncing Dictionary* London: Dent, EPD.
Joos, Martin. 1948. Acoustic phonetics. *Language* 24. 1–136.
Ladefoged, Peter. 1962. *Elements of acoustic phonetics*. Chicago: University of Chicago Press.
Ladefoged, Peter. 1975. *A course in phonetics*. Orlando: Harcourt Brace.
Ladefoged, Peter, J. DeClerk, Mona Lindau, & George Papçun. 1972. An auditory-motor theory of speech production. *Working Papers in Phonetics, UCLA* 22. 48–75.
Li, Min, Chandra Kambhamettu, & Maureen Stone. 2003. EdgeTrak. A program for band-edge extraction and its applications. In *Sixth IASTED International Conference on Computers, Graphics and Imaging*, Honolulu, HI, 82–102.
Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John Ohala & Jeri Jaeger (eds.), *Experimental phonology*, 13–44. Orlando: Academic Press.
Lindblom, Björn, & Johan Sundberg. 1971. Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America* 50. 1166–1179.
Ljung, Lennart. 2010. Perspectives on system identification. *Annual Reviews in Control* 34. 1–12.
MacNeilage, Peter F. 1970. Motor control of serial ordering of speech. *Psychological Review* 77. 182–196.
Maeda, Shinji. 1991. On articulatory and acoustic variabilities. *Journal of Phonetics* 19. 321–331.
Ménard, Lucie, Jean-Luc Schwartz, & Jerome Aubin. 2008. Invariance and variability in the production of the height feature in French vowels. *Speech Communication* 50. 14–28.
Nearey, Terrance M. 1978. *Phonetic features for vowels*. Bloomington: Indiana University Linguistics Club, Indiana.
Noiray, Aude, Khalil Iskarous, Leandro Bolaños, & D. H. Whalen. 2008. Tongue-jaw synergy in vowel height production: Evidence from American English. In *Proceedings of 8th International Speech Production Seminar*, 81–84.
Ostry, David J., Paul L. Gribble, & Vincent L. Gracco. 1996. Coarticulation of jaw movements in speech production: Is context sensitivity in speech kinematics centrally planned? *The Journal of Neuroscience* 16(4). 1570–1579.
Perkell, Joseph S., Frank H. Guenther, Harlan Lane, Melanie L. Matthies, Pascal Perrier, Jennell Vick, Reiner Wilhelms-Tricarico, & Majid Zandipour. 2000. A theory of speech motor control and supporting data from speakers with normal hearing and profound hearing loss. *Journal of Phonetics* 28. 233–272.
Perkell, Joseph S., & Dennis H. Klatt (eds.). 1986. *Invariance and variability in speech processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Perkell, Joseph S., Melanie L. Matthies, Harlan Lane, Frank H. Guenther, Reiner Wilhelms-Tricarico, Jane Wozniak, & Peter Guiod. 1997. Speech motor control: Acoustic goals,

saturation effects, auditory feedback and internal models. *Speech Communication* 22. 227–250.
Perkell, Joseph S., Melanie L. Matthies, Mario A. Svirsky, & Michael I. Jordan. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study. *Journal of the Acoustical Society of America* 93. 2948–2961.
Peterson, Gordon E., & Harold L. Barney. 1952. Control methods used in a study of vowels. *Journal of the Acoustical Society of America* 24. 175–184.
Riley, Michael A., & Michael T. Turvey. 2002. Variability and determinism in motor behavior. *Journal of Motor Behavior* 34. 99–125.
Rost, Gwyneth C., & Bob McMurray. 2010. Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy* 15. 608–635.
Russell, G. O. 1928. *The vowel: Its physiological mechanism as shown by x-ray.* Columbus, OH: Ohio State University Press.
Savariaux, Christophe, Pascal Perrier, & J. P. Orliaguet. 1995. Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *Journal of the Acoustical Society of America* 98. 2466–2474.
Stevens, Kenneth N., & Sheila E. Blumstein. 1978. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 64. 1358–1368.
Stevens, Kenneth N., & Arthur S. House. 1955. Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America* 27. 401–493.
Stevens, Kenneth N., & Arthur S. House. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research* 6. 111–128.
Strange, Winifred, James J. Jenkins, & Thomas L. Johnson. 1983. Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America* 74. 695–705.
Syrdal, Ann K., & H. S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086–1100.
Traunmüller Hartmut. 1981. Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* 69. 1465–1475.
Trubetzkoy, Nikolai. 1969 [1939]. *Principles of phonology*. Berkeley: University of California Press.
Westbury, John, Mary Lindstrom, & Michael McClean. 2002. Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research* 45. 651–662.
Whalen, D. H., Khalil Iskarous, Mark K. Tiede, David J. Ostry, Heike Lehnert-LeHouillier, Eric Vatikiotis-Bateson, & Donald S. Hailey. 2005. HOCUS, the Haskins Optically-Corrected Ultrasound System. *Journal of Speech, Language, and Hearing Research* 48. 543–553.
Whalen, D. H., & Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23. 349–366.
Wood, Sidney. 1975. Tense and lax vowels – degree of constriction or pharyngeal volume. *Lund Working Papers* 11. 109–134.
Wood, Sidney. 1979. A radiographic analysis of constriction location for vowels. *Journal of Phonetics* 7. 25–43.
Wood, Sidney. 1982. X-ray and model studies of vowel articulation. *Lund Working Papers* 23. 1–191.