

# USING OPTICAL FLOW ANALYSIS ON ULTRASOUND OF THE TONGUE TO EXAMINE PHONOLOGICAL RELATIONSHIPS

Kathleen Currie Hall <sup>\*1</sup>, Hanna Smith <sup>1</sup>, Kevin McMullin <sup>2</sup>, Blake Allen <sup>1</sup>, and Noriko Yamane <sup>3</sup>

<sup>1</sup>Department of Linguistics, University of British Columbia, 2613 West Mall, Vancouver, BC V6T 1Z4.

<sup>2</sup>Department of Linguistics, University of Ottawa, 70 Laurier Ave. E., Ottawa, ON K1N 6N5

<sup>3</sup>Graduate School of Integrated Arts and Sciences, Hiroshima University, 1-7-1 Kagamiyama, Higashi-Hiroshima City Hiroshima, Japan 739-8521

---

## Résumé

Cet article examine s'il existe des corrélats articulatoires correspondant aux divers degrés d'une opposition phonologique. On y démontre qu'en anglais, l'amplitude des mouvements impliqués dans l'articulation des voyelles tendues en syllabe ouverte (où elles sont généralement en opposition avec les voyelles relâchées) est supérieure à celle observée en syllabe fermée (où cette opposition est moins marquée). Une analyse de flux optique appliquée à des vidéos échographiques de mouvements de la langue a permis de déterminer l'amplitude de ces mouvements. L'avantage de ce type d'analyse est qu'elle permet une comparaison directe entre les locuteurs et l'obtention de mesures pendant toute la durée d'une production donnée.

**Mots clefs :** échographie, flux optique, opposition, allophony, voyelles

## Abstract

This paper examines whether there are articulatory correlates of differing degrees of phonological contrast. English tense vowels are found to be produced with greater average magnitudes of movement when they occur in closed syllables, where they are generally contrastive with their lax vowel counterparts, than when they occur in open syllables, where they are less contrastive. Magnitude of tongue movement was determined by optical flow analysis of ultrasound videos of tongue movements; optical flow analysis allows for direct comparison of results across speakers and for the extraction of data from the entire timecourse of productions.

**Keywords:** ultrasound, optical flow, contrast, allophony, vowels

---

## 1 Introduction

It is well established that sounds that are contrastive in a given language are often perceived as being more distinct from each other than sounds that are not contrastive in that language, based on reaction times in discrimination tasks and overt similarity rating judgments (e.g., [7], [21], [22], [23], [32]). The conventional wisdom is that these are differences only in the way that sounds are perceived by listeners, rather than reflections of any differences in the way contrastive vs. non-contrastive pairs are produced. Indeed, some studies (e.g., [7]) have found different perceptual results while using acoustically identical stimuli. There is also, however, a small body of evidence that such differences may in fact be encoded acoustically in certain contexts (e.g., [1], [9], [11]). These latter studies share a common result: sounds that are more contrastive in some sense are at least somewhat hyperarticulated relative to their less contrastive counterparts (see §2 for more on quantifying contrastiveness). The results are not entirely conclusive, however. Gick et al. [11] used only one speaker and did not test whether the difference was statistically significant. Goldrick et al. [12] found that the statistically significant results of Baese-Berk and Goldrick [1] hold for only some phonetic distinctions in some phonological contexts (e.g., VOT distinctions are enhanced for contrastive voiceless

stops in initial position, but not for voiced stops). Cristia and Seidl [9] did find consistent differences between phonemic and allophonic pairs of sounds, but found differing results in infant-directed vs. adult-directed speech.

The present paper probes the possibility that there are production differences in regular adult speech that are dependent on the degree of contrast of various sounds. In particular, we examine the possibility of articulatory differences in production using ultrasound imaging. The main research question to be addressed, then, is whether the contrastive status of sounds affects their articulation, with the prediction that contrastive sounds will be articulatorily more distinct than non-contrastive ones. In doing so, we describe the use of optical flow analysis on ultrasound data of tongue movements as a means of extracting time-varying, normalizable data from a relatively large number of participants.

## 2 Degrees of Phonological Contrastiveness

We predicate this study on the assumption that phonological contrastiveness is a gradient phenomenon (e.g., [14], [15], [25]). Two of the primary ways in which contrast is defined are lexical distinction and predictability of distribution, each of which is traditionally treated categorically but can be treated gradiently instead. Typically, lexical distinction is

categorical in the sense that if there is at least one (near) minimal pair that hinges on some pair of sounds, that pair of sounds is deemed to be contrastive. The measure of the *functional load* of a contrast is a gradient instantiation of the same concept: pairs of sounds that distinguish more lexical items have a higher functional load than pairs that distinguish fewer items (see, e.g., [19], [30], [34]). Although there are several methods of calculating functional load, Wedel et al. [34] provide evidence that a simple count of the number of minimal pairs hinging on a contrast (relative to the number of lexical items in a corpus) is an adequate measure, and illustrate its utility in predicting the likelihood of merger: cross-linguistically, pairs of sounds with higher functional loads are less likely to undergo merger than those with lower functional loads.

Traditionally, predictability of distribution is also treated as a categorical parameter: either two sounds are entirely predictably distributed (i.e., in complementary distribution) and are therefore allophonic, or they are not entirely predictably distributed (i.e., there is at least one phonological context in which the occurrence of one vs. the other is not predictable) and are therefore contrastive. Hall [14], however, proposes a gradient measure of predictability of distribution, using the information-theoretic concept of *entropy*, or uncertainty. This measure has been shown to be helpful in documenting phonological changes in progress ([16]), modeling variability in production ([31]), and understanding synchronic phonological harmony patterns ([13]). When applied to two sounds, *a* and *b*, in a phonological relationship, entropy can range between 0 and 1. An entropy of 0 indicates that there is no uncertainty about which of the two sounds occurs in any given context, and is analogous to perfect allophony. An entropy of 1 indicates that *a* and *b* are in perfectly overlapping distributions, and is analogous to perfect contrast.

The sounds of interest for the current study are the tense vowels [i], [u], [o], and [e] in English, which are generally contrastive with their lax vowel counterparts in closed syllables (e.g., there are minimal pairs such as *beat* [i] vs. *bit* [ɪ]; *bayed* [e] vs. *bed* [ɛ]; *who'd* [u] vs. *hood* [ʊ]; *node* [o] vs. *gnawed* [ɔ]). This contrast is largely neutralized in word-final open syllables, however, with only the tense vowels occurring (e.g., *bee* [i] but \*[bi]; *bay* [e] but \*[be]; *who* [u] but \*[hʊ]).<sup>1</sup> Thus, the environments of interest are open vs. closed monosyllabic words; all but one of the stimuli in the experiment were monosyllabic, and using only monosyllables avoids the issue of determining syllable structure in the possible presence of ambisyllabic segments. Both functional load (minimal pair count) and predictability of distribution (entropy) were calculated on a subset of the IPHOD corpus ([30]) containing all and only monosyllabic words of English that have a frequency of occurrence of at

<sup>1</sup> Interestingly, this neutralization occurs for [i]/[ɪ], [u]/[ʊ], and [e]/[ɛ], but not for [o]/[ɔ]; minimal pairs can occur for the latter even in final position (e.g., *know* [o] vs. *gnaw* [ɔ]). This is true even on the assumption of an [ɔ]/[ɑ] merger, in which case the relevant contrast for the current study is [o]/[ɑ]; this will be addressed below.

least one per million using the SUBTLEX frequencies [8] [N = 238 open + 4102 closed = 4340 total uniquely transcribed monosyllables]. The minimum frequency threshold was used to eliminate extremely rare words, such as *thane* or *yaw*, which may not even be known to all speakers, from influencing the calculations; the overall pattern of results is quite similar if such words are included, however. Following [34], homophones were not distinguished (e.g., 'fit' / 'feat' and 'fit' / 'feet' were counted as a single minimal pair). The actual calculations of both functional load and entropy were carried out using the *Phonological CorpusTools* software ([17]). The results are given in Tables 1 and 2.

**Table 1:** Functional load of tense vs. lax vowels in closed and open monosyllables in IPHOD.

Vowel Pair	Functional Load	
	Closed Syllables	Open Syllables
[i] / [ɪ]	98	0
[e] / [ɛ]	86	0
[o] / [ɔ]	41	7
[o] / [ɑ]	56	17
[u] / [ʊ]	8	0

**Table 2:** Predictability of distribution of tense vs. lax vowels in closed and open monosyllables in IPHOD.

Vowel Pair	Pred. of Dist.	
	Closed Syllables	Open Syllables
[i] / [ɪ]	0.95	0.00
[e] / [ɛ]	0.996	0.00
[o] / [ɔ]	0.99	0.67
[o] / [ɑ]	0.97	0.91
[u] / [ʊ]	0.80	0.00

As can be seen, the pair [u] / [ʊ] is distinct from the other pairs in both measures, looking within the set of closed monosyllables. In terms of functional load, there are only 8 minimal pairs hinging on the [u] / [ʊ] distinction, compared to 41–98 pairs for the other three vowels. In this measure, then, this pair is much less contrastive than the other tense-lax pairs. Similarly, there is a much lower entropy value (by 0.15 bits) in closed syllables for the pair [u] / [ʊ] than there is for any of the other three pairs. Both of these measures clearly indicate that the phonological function of this contrast is much weaker than that of the other contrasts: fewer lexical items hinge on this contrast, and if one were given a random closed monosyllabic word of English from a dictionary, it would be easier to guess which of this pair occurs than it would for any of the other three pairs.

The pair [o] / [ɔ] is also distinct from the other three in that it is not non-contrastive in open monosyllables (it has a non-zero functional load and predictability of distribution); rather, it offers an example of a contrast that is simply *weaker* in open syllables than in closed ones. It should be noted, however, that most of the participants in the current

study had, at least impressionistically, an [ɔ] / [ɑ] merger, which is certainly not surprising given the fact that the experiment was conducted in western Canada (see, e.g., Labov et al. [24] : 60). This is not directly a problem, in that the vowels of interest in the study are actually the *tense* vowels, but it does mean that measuring the strength of the relevant tense/lax contrast is somewhat more complicated. Specifically, the lax vowel counterpart of [o] for these speakers may be [ɑ], which means that *all* [ɑ]-containing words must be taken into account and not just those that historically contained [ɔ]. The tables above therefore also show the functional load and predictability of distribution calculations for [o] / [ɑ], under the assumption of an [ɔ] / [ɑ] merger. Including these additional [ɑ] words does not in fact change much about the calculations; this pair is still more contrastive than [u] / [ʊ] and less contrastive than [i] / [ɪ] or [e] / [ɛ] in closed syllables, and is still the only pair of the four that is contrastive in open syllables. The primary difference is that the magnitude of the difference in the contrast between closed and open syllables is much smaller if one assumes that there is a merger. That is, while it is still the case that [o] / [ɑ] is less contrastive in open syllables than closed syllables, the two environments are more similar to each other in the merged data than they are in the unmerged data, especially with respect to predictability of distribution.

We now turn to the ultrasound study used to examine whether these differences in contrastiveness have articulatory consequences.

### 3 Methodology

Stimuli consisted of 78 English target words with tense vowels in stressed word-final syllables. All but one of these words were in fact monosyllabic; the one exception was the word *delay*, which has [e] in a final stressed open syllable. There were 10 closed-syllable words for each of [i], [e], [u], and [o], and 10 open-syllable words for each of [i], [e], and [o], plus 8 open-syllable words for [u]. Additionally, there were 46 filler words with lax vowels in stressed word-final syllables, all monosyllabic. Within these, there were 10 words with each of [ɪ] and [ɔ] in closed syllables; 11 with [ɛ] in closed syllables; 8 with [ʊ] in closed syllables; and 7 with [ɔ] in open syllables. All stimuli are presented in Appendix A.

Twenty-four female speakers participated in the study. It has been suggested (Eric Vatikiotis-Bateson, p.c.) that ultrasound imaging might be clearer for female rather than male speakers because of the generally higher degree of calcification in males as compared to females (e.g. [27]). Ten of the 24 participants were excluded from analysis either because of evidence that they were non-native speakers of standard North American English (e.g., having grown up outside of North America or reporting an alternative first language) and/or because of technical difficulties during recordings. This left a total of 14 participants, who were between the ages of 18 and 26, with an average age of 21.5. Participants were paid \$20 each for their participation. No included participants reported any

speech or hearing difficulties.

Participants were tested one at a time. They were seated in a fixed chair with a headrest to help minimize movement during the experiment while still allowing for natural productions. An Aloka SSD-5000 ultrasound machine was used to collect ultrasound. A UST-9118 endovaginal 180° electronic curved array probe was placed firmly under the participant's chin. The probe was positioned roughly halfway between the chin and the neck, at approximately the midline (sagittal) position, at an approximately 90° angle to the floor (all aspects judged by two experimenters, viewing from both the front and the side). Slight adjustments to the probe position and pressure under the chin were made to ensure the ultrasound image captured the entire tongue and was as clear as possible. After this point, participants were asked to be as still as possible during the recording. The probe was held with a mechanical arm, which was connected to a pole adjacent to the chair, with a layer of ultrasound gel between the probe and the skin. Two-dimensional mid-sagittal ultrasound video recordings of the tongue were recorded digitally directly to an attached computer at a rate of 30 frames / second.

Productions were simultaneously audio-recorded onto the computer recording the ultrasound data, using a Shure SM63LB Dynamic handheld microphone placed in a floor-stand approximately 18 inches from the participant's mouth. Both the audio and video recordings were made in iMovie.<sup>2</sup>

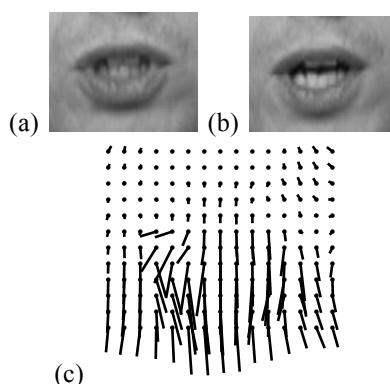
A laptop computer was placed at a comfortable viewing distance in front of the participant. Stimuli were presented one word at a time on the screen, with one of the experimenters advancing to the next word after it had been produced by the participant. The 124 total stimuli were presented one time through, in random order, though it should be noted that the first four participants (all of whom were included in the final data analysis) happened to see the words in the *same* random order as each other.

### 4 Analysis techniques

The ultrasound video images were subjected to optical flow analysis (OFA; e.g., [3], [10], [18], [20], [26]), using *FlowAnalyzer* software developed by Barbosa [2], which uses the implementation of OFA described in [3]. OFA provides a way of measuring apparent motion by comparing the difference in brightness of individual pixels from frame to frame.

Consider Figure 1, from Fleet and Weiss ([10]: 10). Figures (1a) and (1b) show two adjacent frames in a video, where the lips have progressed from being more closed to more open. Figure (1c) shows the optical flow field associated with this frame sequence; each pixel is associated with a vector showing apparent motion between frames. This example, of course, illustrates using OFA on direct video of the articulators; in the current study, we applied OFA to ultrasound videos of the tongue rather than video of the tongue itself.

<sup>2</sup> iMovie '11 (vers. 9.0.9), available from Apple Inc., running on Mac OS X 10.10.5.



**Figure 1:** Optical flow field (c) resulting from the apparent motion between adjacent video frames (a)-(b), from Fleet and Weiss ([10]: 10).

OFA has several advantages over standard measures of articulatory posture, especially for the purposes of the present research question. OFA allows for the extraction of information from the entire production of a sound or word, rather than using still images from pre-designated timepoints within the production. Thus, for productions where there is no *a priori* reason to suspect that differences would be localized to particular temporal regions, OFA permits researchers to look for differences throughout. It is also possible to obtain measurements from different physical regions of the video (e.g., isolating the tongue tip, body, or root) separately, to examine effects on these various regions independently, though one can also examine the video as a whole. Furthermore, OFA is relatively fast and automatic (see also [26]). While it may still be necessary to annotate accompanying sound files in order to determine the timepoints of particular intervals of interest, OFA drastically reduces the overall amount of time needed to analyze ultrasound data. Indeed, it makes it possible to analyze ultrasound video data with roughly the same efficiency as acoustic data. Finally, OFA allows for direct comparison of measurements across speakers, which is often not the case for articulatory posture data (though cf. [35] for an example of normalization across tongue curves). OFA data can easily be normalized within a speaker, using, for example, a standard z-score normalization, which then allows data to be pooled across participants.

In order to analyse the data in this study, the audio was first extracted to .wav files from the video recordings of each speaker, using a Python script.<sup>3</sup> The target vowel in each word was identified and delimited using a *Praat* TextGrid ([6]). Vowel boundaries were identified by looking for clearly visible formant structure and increased

<sup>3</sup> Specifically, conversion was done using the `convert_mov_to_wav.py` script in the Ultrasound Analysis package available (March 2017) here : <https://github.com/bhallen/ultrasound-analysis/>, which in turn makes use of the FFmpeg software, available (March 2017) at <http://www.ffmpeg.org/>.

intensity as compared to the surrounding sounds. Interval boundaries were placed at zero-crossings of the waveform.<sup>4</sup>

*FlowAnalyzer* was used to extract OFA information from the complete ultrasound video files. No particular regions were specified; movement from all regions of the tongue were included (i.e., movement from the entire video image), as there was no *a priori* expectation that any regions of the tongue would be more likely than others to demonstrate differences based on contrastive status; this is precisely one of the reasons that this type of generalized OFA is advantageous as compared to either edge-tracking analyses or even other types of OFA such as that used in [26]. As described in detail in [3], *FlowAnalyzer* reduces the high-dimensionality of a full optical flow field (with a separate measurement for each pixel in an array) to a single dimension by summing the magnitudes of movements of all of the pixels between a single pair of frames (a ‘frame-step’) to result in one total magnitude measure for that frame-step, given in number of pixels moved, as shown in (1).

(1) Single magnitude measure for the  $n$ -th region of interest at time  $k$ , where  $\langle\langle\| \cdot \| \rangle\rangle$  denotes the vector magnitude, and  $x_i, x_f, y_i, y_f$  are the initial and final boundary positions of the region of interest in the horizontal and vertical directions, respectively » (Barbosa et al. [3] : p. 174, eq. 2)

$$v_n(k) = \sum_{x=x_i}^{x_f} \sum_{y=y_i}^{y_f} \|\vec{v}(x, y, k)\|$$

This number is then divided by the number of pixels in the given region of interest, to result in a mean magnitude of movement, in pixels, for a given frame-step.

Note that the measure « magnitude-per-frame-step » is not directly a measure of magnitude of movement (i.e., a distance measure); it is instead a measure of the *rate* of movement, being a measure of distance (magnitude, i.e., number of pixels) per unit time, where the time is one frame-step.<sup>5</sup>

<sup>4</sup> Note that unfortunately, the quality of the acoustic recordings accompanying the ultrasound videos in this study is not particularly good. Recordings were made in an open room rather than a sound-attenuated booth, with the microphone relatively far away from the participants, and there was a fair bit of background noise. While the recordings were good enough to allow for rough delimitation of the edges of vowels (and with a frame rate in the video of only 30 frames per second, a high level of resolution isn’t needed), more fine-grained analysis (e.g., of the formant structure of vowels in open vs. closed syllables) is not possible.

<sup>5</sup> We are deeply grateful to an anonymous reviewer for extensive discussion of this point and its implications. This should not be confused with saying that these are the velocities within a given frame; that would be obtained by multiplying the magnitude in the frame-step by the frame rate (30 fps), to result in a measure of how fast the pixels were moving in a particular frame, in pixels per second.

The output of the software is a table of values, one per frame in the video, giving the timestamp of the frame along with the mean magnitude of movement in the x- and y-dimensions for that frame-step, along with the mean total magnitude measure for the frame-step. (Note that these are indeed *magnitude* measures per frame-step and do not include information about the directionality of movement—that is, an upward movement followed by a downward movement of the same distance will have twice the magnitude, rather than having a measure of zero.)

This dimensionality reduction is different from other implementations of optical flow analysis (e.g., [26]), which generally maintain more of the details within the field. As Barbosa et al. [3]: 174 explain, however, «the temporal variation of this seemingly impoverished measure is surprisingly well-coordinated with time-varying measures made in other domains (for example, the RMS amplitude of the speech acoustics)»; see also discussion in [4]. For the present purposes, the reduction is particularly advantageous because the question is really whether there is a correlation between the magnitude of the phonological contrast and the magnitude of tongue movement as a whole, so having a single dimension for each side of the correlation is beneficial. If one wanted to know more specifics about the *mechanics* of the movement and especially about either the directionality or the differences across different regions of the tongue, then a less reductionist approach would be preferable.

Returning to the current analysis, each vowel consists of some (differing) number of frames, but each frame-step encapsulates the same duration from one frame to the following frame. In order to get the total magnitude of tongue movement in a particular vowel gesture, then, the mean magnitudes per frame-step must be summed over all the frame-steps in the vowel. This summation allows one to look for a direct correlation between the magnitude of the phonological contrast and the magnitude of tongue movement.

The drawback of this summative approach from an analysis perspective is that longer vowels have more frames that go into the calculation of total magnitudes, and could possibly show greater magnitudes simply because of a longer duration for reasons other than the degree of their phonological contrastiveness.

To unpack this, consider the experimental hypotheses. Under the null hypothesis that tense vowels are articulatorily the *same* in contrastive and non-contrastive positions, there are two primary possibilities for how this «sameness» could manifest itself in a way measurable by OFA: either the total magnitudes could be the same, or the magnitudes-per-frame-step could be the same. In the former case, longer vowels would show equal total magnitudes as shorter vowels, but would therefore have to have smaller magnitudes per frame-step to compensate. In the latter case, longer vowels would have the same magnitudes-per-frame-step as shorter vowels, but would therefore end up with larger total magnitudes. The logical alternative hypotheses here are that tense vowels in contrastive positions have greater total magnitudes than those in less-contrastive

positions, but not simply because they are longer; or that vowels in contrastive positions have greater magnitudes-per-frame-step, but not simply because they are shorter.

Given that in the current dataset, the contrastive vowels are in closed syllables, they are independently shorter on average than their open-syllable, non-contrastive (or less-contrastive) counterparts. Specifically, vowels in closed syllables were an average of 7.72 frames long, while those in open syllables averaged 9.85 frames, which is statistically significantly longer [ $t(970.6) = 13.58, p < 0.001$ ]. Similar statistically significant differences are found for each vowel quality individually. This difference in duration makes it impractical to directly compare either total magnitudes or magnitudes per frame-step, as differences could be attributed to durational differences rather than contrastive status.

To test for the effects of contrastive status, then, we run linear mixed-effects regressions in which both duration and contrastive status will be used to predict total magnitude of movement.<sup>6</sup> By first showing that these two predictors are not collinear with each other, and then showing that each has a statistically significant effect on magnitude, we conclude that the total magnitude of tongue movement is dependent on the contrastive status of the vowel's position.

*Praat* TextGrids were used to determine the time stamps of the beginning and end points of each of the target vowels. These frames were then extracted from the output of the OFA data, giving a list of mean magnitude of movement per frame-step for each frame contained within a target vowel, for each speaker. Only the total mean magnitudes for each frame-step were included, not individual horizontal and vertical magnitudes.

It is quite likely that individual speakers vary widely in their actual movements during production, given different anatomy and speech styles. Thus, the per-frame-step magnitude data for each speaker was subjected to a z-score normalization, such that the mean magnitude of movement per frame-step across all vowels for each speaker was set to 0, with a standard deviation of 1. To then calculate the total magnitude of movement in any particular vowel, the normalized values for all frame-steps in that vowel were summed. This normalization allows for direct comparison of data across speakers.

## 5 Results

The summed normalized magnitude of movement data for each of the four tense vowels in closed vs. open syllables is shown in Figure 2. There is one (summed) measurement per word per speaker in each box, e.g., 10 words \* 14 speakers = 140 tokens for [i] in open syllables. Outliers of more than three standard deviations from the mean total for a given vowel were removed; there was one such outlier for [e] and

---

<sup>6</sup> We note that we did also do the analyses on the magnitude-per-frame-step measures as well, with similar global results, i.e., in both cases, we find a statistically significant effect of contrastive status separate from that of duration, in the phonologically expected manner.

two each for [o] and [u]. As can be seen, the total normalized magnitude of movement is greater in closed syllables than in open syllables for [i], [e], and [o].

This result must be interpreted carefully, though, as discussed above. The greater values for normalized total magnitude in closed syllables could, for example, be caused by having larger magnitudes per frame-step in an effort to have equal total magnitudes in a vowel that has fewer frames. Thus, a simple comparison of the magnitudes is not sufficient to show that syllable type matters beyond duration.

Linear mixed-effects models in R ([5], [29]) predicting the total magnitude of movement for a given vowel type ([i], [e], [o], or [u]) from the fixed effects of duration and syllable type (open vs. closed), with random intercepts for participant and word, and random by-participant and by-item slopes for the effect of syllable type, however, do indicate that syllable type plays a significant role in its own right. (Note that details of the models are given in Appendix B; relevant aspects are reported in the text.)

First, for each model, we examine the collinearity of the two predictor variables, duration and syllable type, by computing the variance inflation factor (VIF), to ensure that they are not simply duplicating each other. VIFs around a value of 1 indicate that two predictors are not particularly correlated, while those greater than a threshold of 5 or sometimes 10 are considered problematically correlated (see discussion in [28]). In the current situation, the VIFs ranged from 1.21 for [o] to 1.44 for [i], indicating that syllable type and duration are not particularly correlated.

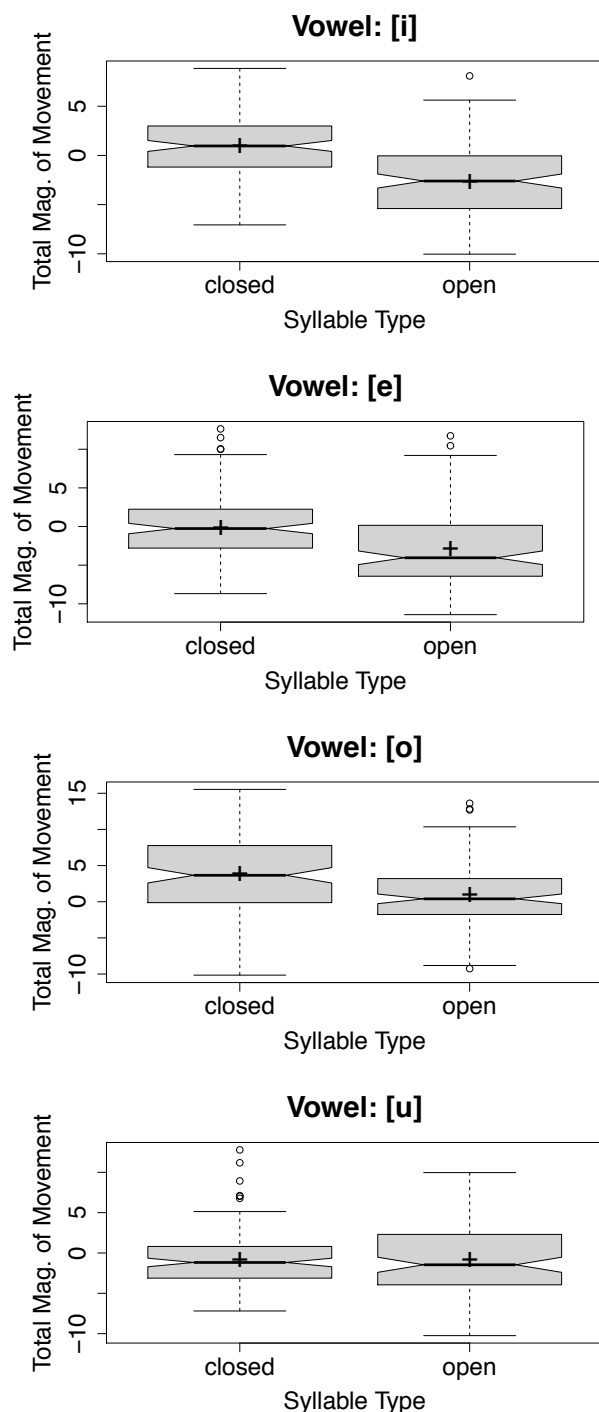
Second, the baseline for each model was taken to be the total magnitude value in closed syllables; thus, if the hypothesis that vowels have smaller magnitudes of movement when they are in less-contrastive positions is correct, then we expect to see statistically significant *negative* estimates for open syllables. Given that open syllables are also longer than closed syllables, we might also see that duration has a negative estimate, so that longer vowels consist, on average, of smaller individual movements.

Finally, for each model, the statistical significance of syllable type was determined by performing a likelihood ratio test of the model in question to a model that was equivalent except that syllable type was not included as an effect. In all cases, visual inspection of residual plots also indicated that the standard assumptions of homoscedasticity and normality for linear models were met.

For all models, the effect of duration was indeed in the expected direction (negative) and was statistically significant or nearly so (in the case of [u];  $p = 0.07$ ); the details of the results for duration in each model are not further given in the text, as the focus here is the examination of whether syllable type *also* matters.

For [i], syllable type significantly affected total magnitude ( $X^2(5) = 22.33$ ,  $p < 0.001$ ), as predicted, with open syllables reducing the overall magnitude by about 2.41 standardized units,  $\pm 0.91$  (standard errors). Given that this model had random by-word and by-participant slopes for the effect of syllable type, we can also examine the

individual estimates for the effect of open syllables for each word and each participant. In this case, all 21 [i]-containing words and 13 of the 14 participants were assigned negative estimates for open syllables, further confirming the hypothesis.



**Figure 2:** Normalized, summed magnitude of movement data for tense vowels in closed vs. open syllables. Horizontal lines within each plot show the median values; plus signs indicate the mean values.

Similar results hold for [e]: syllable type significantly affected total magnitude ( $X^2(5) = 21.43$ ,  $p < 0.001$ ), as predicted, with open syllables reducing the overall magnitude by about 1.54 standardized units,  $\pm 1.25$  (standard errors). There is slightly less uniformity across words and participants for this vowel, although the trend is the same: 18 of the 21 [e]-containing words, and 10 of the 14 participants, were assigned negative estimates for open syllables.

For [o], the results are similar, but do not quite reach statistical significance under the assumption of  $\alpha = 0.05$ . Syllable type tended to affect total magnitude ( $X^2(5) = 9.99$ ,  $p = 0.076$ ) in the direction predicted, with open syllables reducing the overall magnitude by about 2.16 standardized units,  $\pm 1.20$  (standard errors). In terms of individual words and participants, 20 of the 22 [o]-containing words and 13 of the 14 participants were assigned negative estimates for open syllables.

The results for [u], however, are decidedly different, as can be seen visually in Figure 2. The effect of syllable type on total magnitude was not close to being significant ( $X^2(5) = 1.42$ ,  $p = 0.92$ ), and the estimate was in the opposite direction (i.e., positive). Indeed, 15 of the 19 [u]-containing words, and all 14 participants, were assigned positive estimates for open syllables.

These results indicate that there tends to be greater total magnitude of movement of tense vowels in closed syllables, where there is a greater potential for lexical contrast, than in open syllables, where the potential is smaller, beyond simply the effect of duration. This is the case for [i] and [e], where the difference between closed and open syllables is categorical, and also for [o], where the difference between contrastiveness in closed and open syllables is simply one of degree. These results will be discussed in more detail in §6, as will the lack of an effect for [u].

First, though, it should be noted that, while the phonetic contexts were not controlled for in this experiment, post-hoc examination of a subset of the stimuli that are matched phonetically suggest that these results are not driven exclusively by context-specific articulations. A post-hoc comparison group was created, containing only closed-syllable, bilabial-final<sup>7</sup> words for which there were open-syllable counterparts with closely matched onsets. The following words were included in this « matched » subset: *beam*, *bee*, *team*, *tea*, *hoop*, *who*, *tube*, *two*, *slope*, *low*, *babe*, and *bay*.

The statistical results for this matched subset were mixed. Because there were so few items, random slopes by syllable type were not possible, and only random intercepts were included. The estimates for open syllables for both [i] and [e] in this subset were negative, but not quite statistically significant ( $X^2(1) = 1.92$ ,  $p = 0.16$  for [i] and  $X^2(1) = 3.48$ ,  $p = 0.06$  for [e]). The effect for [o] disappeared entirely, with the estimate being positive and not close to statistically significant ( $X^2(1) = 0.90$ ,  $p = 0.34$ ).

<sup>7</sup> Bilabial-final words were chosen to minimize co-articulatory effects on tongue movement between the vowel and the coda consonant.

The effect for [u] was similarly not significant, though interestingly, the estimate here was in fact negative and the result trended toward significance ( $X^2(1) = 2.86$ ,  $p = 0.09$ ).

Future testing with larger datasets that are phonetically matched will need to be done to truly understand the role of phonetic context. At the same time, the results for [i] and [e] in particular seem to be consistent regardless of context, with vowels in open syllables displaying smaller total normalized magnitudes of movement as compared to their closed-syllable counterparts. Given the weaker nature of the contrasts for both [o] and [u], discussed in §6 below, this suggests that there does need to be a clear-cut contrast phonologically in order for an articulatory effect to be present.

## 6 Discussion and Conclusions

The above results strongly suggest that there is an articulatory difference between most English tense vowels when produced in closed vs. open syllables. The lack of an effect for [u] suggests two things. First, whatever *is* causing the difference for the other vowels, it is unlikely to be the simple fact of syllable structure itself. That is, it doesn't seem to be the case that closed syllables simply involve larger magnitudes of tongue movement than open syllables, regardless of phonological contrastiveness. Second, there seems to be some critical degree of contrast that is relevant. Given that functional load and type-based entropy largely pattern together when it comes to distinguishing [u] / [ʊ] from the other pairs, it is not possible from this study alone to determine whether one of these is in any sense the “critical” factor, and if so, what the critical aspect of that factor might be.

One can speculate, however, that there is some threshold value above which articulations are hyperarticulated relative to other contexts, presumably because they are deemed “contrastive enough” to be relevant. For entropy, this threshold would need to be somewhere between 0.67 bits (the entropy of [o]/[ɔ] in open syllables, where relative hyperarticulation doesn't occur) and 0.95 bits (the entropy of [i]/[ɪ] in closed syllables, the lowest entropy at which the relative hyperarticulation does occur). Under the assumption of an [ɔ] / [ɑ] merger, however, the window for the threshold is quite narrow, as it would need to be somewhere between 0.91 and 0.95. For functional load, there may be some minimum number of minimal pairs, greater than 8 (the number of pairs for [u]/[ʊ], where there is no effect) and smaller than 41 (the number for [o]/[ɔ], where there is one), required for relative hyperarticulation to take place. Under the assumption of an [ɔ] / [ɑ] merger, the interval would be between 17 and 56.

The current study is not fine-grained enough to tease the functional load and entropy measures apart, though it does seem somewhat more plausible that a threshold could be found in the intervals defined by minimal pairs than by predictability of distribution, at least under the assumption of an [ɔ] / [ɑ] merger. Nor can it eliminate other possibilities, such as the generally low frequency (both lexically and in use) of [u]/[ʊ] as compared to the other



vowel pairs (which is, of course, correlated with the measures used here). At the same time, it does show clear evidence that some measure of contrastiveness is correlated with articulation, with sounds that are *more* lexically contrastive being hyperarticulated relative to sounds that are *less* lexically contrastive; more specifically, sounds that are more contrastive involve larger average movements within any given frame. This correlation is true, however, only if one accepts claims that phonological contrast is not a binary notion but rather a gradient one. Finally, the current study has illustrated the utility of optical flow analysis in the study of ultrasound data. While OFA does not directly reveal patterns of tongue posture, it can tell us about what the tongue is doing continuously during articulation, and the resulting measures can be normalized and directly compared across participants.

Though the current study raises a number of questions – what kind and degree of contrastiveness matters for affecting articulations? is contrastiveness the causal factor, or is it simply also correlated with the causal factor? is this a case of hyperarticulation of contrasts or hypoarticulation of non-contrasts? are there acoustic consequences of these differences? what exactly is the role of phonetic context in determining magnitude of movement? – it is our hope that the methodology and initial results reported here will indeed spur further research that can answer these questions.

## References

- [1] M. Baese-Berk and M. Goldrick. Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24 :527, 2009.
- [2] A. V. Barbosa. FlowAnalyser. [Computer Program], 2013. Available online : [https://www.cefa.org/~adriano/optical\\_flow/](https://www.cefa.org/~adriano/optical_flow/)
- [3] A. V. Barbosa, H. C. Yehia, and E. Vatikiotis-Bateson. Linguistically valid movement behavior measured non-invasively. In R. Göcke, P. Lucey, and S. Lucey (eds.), *Auditory and visual speech processing – AVSP08*. 173, 2008.
- [4] A. V. Barbosa, H. C. Yehia, and E. Vatikiotis-Bateson. Temporal characterization of audio-visual coupling in speech. In *Proceedings of Meetings on Acoustics*, 1, 2008.
- [5] D. Bates, M. Maechler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67 :1, 2015.
- [6] P. Boersma and D. Weenink. Praat: a system for doing phonetics by computer. Available from [www.praat.org](http://www.praat.org). 1992–2015.
- [7] A. Boomershine, K. C. Hall, E. Hume, and K. Johnson. The influence of allophony vs. contrast on perception: The case of Spanish and English. In P. Avery, B.E. Dresher, and K. Rice (eds.), *Contrast in phonology: Perception and acquisition*. 145, 2008.
- [8] M. Brysbaert and B. New. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41, 2009.
- [9] A. Cristia and A. Seidl. The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, 41 :1, 2013.
- [10] D. J. Fleet and Y. Weiss. Optical Flow Estimation. In N. Paradigios, Y. Chen, and O. D. Faugeras (eds.), *Handbook of Mathematical Models in Computer Vision*. 239; 2006.
- [11] B. Gick, D. Pulleyblank, F. Campbell, and N. Mutaka. Low vowels and ATR harmony in Kinande. *Phonology*, 23 :1, 2006.
- [12] M. Goldrick, C. Vaughn, and A. Murphy. The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America*, 134 :EL172, 2013.
- [13] D. C. Hall and K. C. Hall. Marginal contrasts and the Contrastivist Hypothesis. Paper presented to the Linguistics Association of Great Britain, London, 2013.
- [14] K. C. Hall. A probabilistic model of phonological relationships from contrast to allophony. Columbus, OH: The Ohio State University Doctoral dissertation, 2009.
- [15] K. C. Hall. A typology of intermediate phonological relationships. *The Linguistic Review*, 30 :215, 2013.
- [16] K.C. Hall. Documenting phonological change: A comparison of two Japanese phonemic splits. In S. Luo (ed.), *Proceedings of the 2013 Annual Meeting of the Canadian Linguistic Association*. Toronto: Canadian Linguistic Association, 2013. Available online: <http://homes.chass.utoronto.ca/~cla-acl/actes2013/actes2013.html>.
- [17] K. C. Hall, B. Allen, M. Fry, S. Mackie, and M. McAuliffe. *Phonological CorpusTools*, Version 1.0. [Computer program], 2015. Available from <https://sourceforge.net/projects/phonologicalcorpustools/>.
- [18] K. C. Hall, C. Allen, K. McMullin, V. Letawksy, and A. Turner. Measuring magnitude of tongue movement for vowel height and backness. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow. Paper #0854.
- [19] C. F. Hockett. The quantification of functional load: A linguistic problem. *U.S. Air Force Memorandum RM-5168-PR*, 1966.
- [20] B. K. P. Horn, B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17 :185, 1981.
- [21] E. Hume and K. Johnson. The impact of partial phonological contrast on speech perception. In M.-J. Solé, D. Recansens, and J. Romero (eds.), *Proceedings of the Fifteenth International Congress of Phonetic Sciences*. 2385, 2003.
- [22] J. J. Jaeger. Testing the psychological reality of phonemes. *Language and Speech* 23 :233, 1980
- [23] N. Kazanina, C. Phillips, and W. J. Idsardi. The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America* 103 :11381, 2006.
- [24] W. Labov, S. Ash, and C. Boberg (eds.). *Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin : Mouton de Gruyter.
- [25] D. R. Ladd. "Distinctive phones" in surface representation. In L. M. Goldstein, D. H. Whalen, and C. T. Best (eds.), *Laboratory Phonology 8* :3, 2006.
- [26] S. Moisisik, H. Lin, and J. H. Esling. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association* 44 :21, 2014.
- [27] M. Mupparapu and A. Vuppapapati. Ossification of laryngeal cartilages on lateral cephalometric radiographs. *Angle Orthodontist*, 75, 2005.
- [28] R. O'Brien. A caution regarding rules of thumb for Variance Inflation Factors. *Quality & Quantity* 41 :673, 2007.
- [29] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>. 2015.
- [30] D. Surendran and P. Niyogi. Measuring the functional load of phonological contrasts. In *Tech. Rep. No. TR-2003-12*, 2003.
- [31] P. Thakur. Sibilants in Gujarati phonology. Paper presented at the workshop on Information-theoretic approaches to linguistics, University of Colorado – Boulder, 2011.
- [32] N. S. Trubetzkoy. *Principles of Phonology*. Berkeley: University of California Press. 1969 [1939].



[33] K. I. Vaden, H. R. Halpin, and G. S. Hickok. *Irvine Phonotactic Online Dictionary*, Version 2.0. Available from: <http://www.iphod.com>, 2009.

[34] A. Wedel, A. Kaplan, and S. Jackson. High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128 :179, 2013.

[35] N. Zharkova, N. Hewlett & W. J. Hardcastle. Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control* 15 :118, 2011.

## Appendix A

		Tense Vowels (Target Words)			
		[i]	[e]	[o]	[u]
<b>Closed Syllables</b>	bead	babe	boat	bood	
	beam	bayed	bode	boot	
	cheek	cake	code	doom	
	feast	cave	cove	duke	
	leaf	face	foam	dune	
	meat	fame	ghost	food	
	seed	gate	globe	fool	
	sheep	mail	goal	hoop	
	team	phase	nose	moose	
	teen	safe	slope	suit	
<b>Open Syllables</b>	bee	bay	blow	blue	
	fee	clay	bow / go <sup>8</sup>	chew	
	glee	day	doe	clue	
	key	delay	flow	coo	
	knee	hay	hoe	stew	
	me	jay	Joe	two	
	plea	may	low	who	
	spree	ray	mow	zoo	
	tea	stay	toe		
	tree	way	woe		

## Lax Vowels (Filler Words)

		[ɪ]	[ɛ]	[ɔ]	[ʊ]
<b>Closed Syllables</b>	bid	bed	boss	foot	
	bin	bell	cob	full	
	dish	chef	cough	good	
	fib	head	dot	hood	
	gill	jet	job	pull	
	hip	mesh	moss	put	
	kid	mess	pause	soot / could <sup>4</sup>	
	kiss	pep	pod	wood	
	pig	pet	pot		
	pit	test	top		
<b>Open Syllables</b>		web			
			bawdy		
			body		
			claw		
			flaw		
			jaw		
		law			
		paw			

## Appendix B

### Linear Mixed-Effect Models, Full Dataset

		Fixed effect	Estimate	Standard error	t-value
<b>[i]</b>	(Intercept)		4.79	1.16	4.14
	Open syll.		-2.41	0.91	-2.64
	Duration		-0.54	0.14	-3.83
<b>[e]</b>	(Intercept)		4.12	1.15	3.57
	Open syll.		-1.54	1.25	-1.23
	Duration		-0.53	0.12	-4.51
<b>[o]</b>	(Intercept)		7.43	1.75	4.25
	Open syll.		-2.16	1.20	-1.80
	Duration		-0.42	0.18	-2.37
<b>[u]</b>	(Intercept)		0.99	1.19	0.83
	Open syll.		0.61	1.07	0.57
	Duration		-0.26	0.14	-1.84

<sup>8</sup> The first four participants were run with the words “bow” [bo] and “soot” [sɔt]. There were consistent errors in production of these words, as [baʊ] and [sʊt], probably due to ambiguity in the former case and unfamiliarity in the latter case. Hence, these words were replaced with “go” and “could,” respectively, for the remaining participants.

**Linear Mixed-Effect Models, Matched Dataset**


	Fixed effect	Estimate	Standard error	t-value
[i]	(Intercept)	9.91	2.94	3.37
	Open syll.	-1.60	1.14	-1.40
	Duration	-1.22	0.35	-3.52
[e]	(Intercept)	1.27	2.54	0.50
	Open syll.	-2.52	1.17	-2.15
	Duration	-0.34	0.26	-1.30
[o]	(Intercept)	7.07	3.29	2.15
	Open syll.	2.16	2.13	1.02
	Duration	-0.48	0.48	-1.00
[u]	(Intercept)	-0.36	1.29	-0.28
	Open syll.	-1.49	0.86	-1.74
	Duration	-0.14	0.18	-0.77



**HGC ENGINEERING**

- > Noise & Vibration Control in Land-use Planning
- > Noise & Vibration Studies: Residential and Commercial
- > Building Acoustics, Noise & Vibration Control
- > Land-use Compatibility Assessments
- > Third-party Review of Peer Reports
- > Expert Witness Testimony

905-826-4546  
 answers@hgcengineering.com  
[www.hgcengineering.com](http://www.hgcengineering.com)



**INDOOR OUTDOOR NOISE CONTROL**

Point, Line\* and Surface\* sources  
 Sound Transmission tools  
 Simulate - Measure STI, RT and other Parameters

Import any geometry easily in .dxf format  
 SketchUp plugin included

User-friendly interface  
 A wealth of graphics for your reports

Available in Basics, Industrial\*, Auditorium and Combined\*

IEC 60268 • STI, ISO 14257 • Workplaces  
 ISO 3382-3 • Open Plan Offices



**www.odeon.dk**  
 Measurements - Simulations - Auralisation