Measuring Magnitude of Tongue Movement for Vowel Height and Backness

Kathleen Currie Hall, Claire Allen, Kevin McMullin, Veronica Letawsky, Alannah Turner

University of British Columbia
kathleen.hall@ubc.ca

ABSTRACT

Optical Flow Analysis (OFA) has recently been introduced as a fast, easy, and reliable means of extracting articulatory information from ultrasound video ([2], [3], [6], [9]). This paper illustrates how one primary measurement that is extracted using OFA, magnitude of movement, correlates with the standard phonological vowel categories of height and backness. By establishing a baseline for these measures, we pave the way for future studies to examine how magnitudes change in particular phonological, social, or other contexts.

Keywords: Optical Flow Analysis, ultrasound, magnitude, articulation, vowels

1. INTRODUCTION

Optical Flow Analysis (OFA; [7], [5]) is a technique used to extract apparent movement from video data by comparing the difference in brightness of individual pixels from frame to frame. It has relatively recently been introduced into the toolbox of linguists [2], [3], [6], [9]. In particular, [2], [6], and [9] have suggested a great potential for OFA in analyzing articulatory ultrasound video data, because of the ability to (1) easily extract data from all video frames, rather than single frames as is common with static postural analyses, (2) normalize the resulting data across participants, and (3) extract data quickly, objectively, and relatively effortlessly from extended videos.

While the advent of such an analysis technique is promising, it is currently difficult to evaluate the results of studies that make use of it, simply because there are no baseline measures of what kinds of values are expected from this analysis. The technique results in frame-by-frame measures of the apparent magnitude of movement (MM) of objects (such as the surface of the tongue) in a video. The current study seeks to establish what the magnitude correlates are for the height and backness classifications of vowel categories. While we anticipate that there may be further subtle distinctions that can be made by examining individual sections of the tongue (tip, blade, root), and that the MMs will also vary by the contexts in which vowels are uttered, it is precisely in order to understand these further distinctions that the current study is crucial.

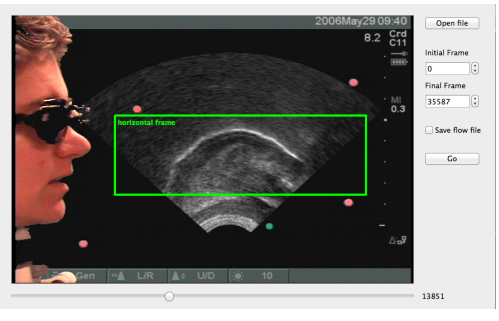
2. PROCEDURE

The input was the ultrasound video files for two male subjects (subjects 1 and 2) used in [8];[1] both are native speakers of American English with phonetic training. As described there (147), "The tongue was imaged using a SonoSite TITAN portable ultrasound machine with a C-11/7-4 11-mm broadband curved array transducer placed under the chin, generating a mid-sagittal section from near the tongue root to near the tongue tip at a rate of 28 scans per second, output as 29.97 fps analog video." Note that this does result in a slight overlap between frames, such that each frame includes approximately 3 ms worth of the scanning from the previous frame.

The videos were subjected to OFA using FlowAnalyzer [1], which uses the algorithms defined by [7]. The input is a video file, and the output is a spreadsheet with estimates for the horizontal and vertical magnitudes of movement for each frame, along with a total magnitude measure, which is the sum of all the Euclidean magnitude measures for each pixel. Specific regions of interest or disinterest can be specified; for example, measurements can be included only for areas that include the tongue tip or tongue body, if desired.

Because the original ultrasound video files had a regular video of the speaker superimposed on them, along with video tracking dots that were on the glasses worn by the speakers (see Figure 1), a horizontal window for OFA was selected that excluded the superimposed video images, but included the bulk of the tongue movement. Note that

Figure 1: Example horizontal frame for OFA using FlowAnalyzer [1].



all analyses discussed in this paper were also repeated with a vertical frame that included a greater vertical range and a smaller horizontal range; the results were virtually identical.

All frames in the video were analyzed for the apparent magnitude of movement from the previous frame. The frames corresponding to vowels were selected by comparing the timestamps of the frames with the timestamps of Praat TextGrids [4] that delimited the edges of the segments produced (again, the TextGrids came from [8]).

The stimuli were nonsense words of the form [V_1XV_2], where $V_1 = V_2$ and were from the set {[i], [ɑ], [u]} (we refer to these as the "flanker" vowels), and X was a "target" phone. The target phones in [8] included both consonants and vowels, but only vowels are considered here. In the current study, we examine MMs in both flanker and target vowels.

All vowels were labeled according to standard classifications for height and backness. The number of tokens for each vowel are given in Table 1.

Table 1: Counts of vowel tokens by context. Height labels: L(ow), M(id), H(igh); Backness labels: F(ront), C(entral), B(ack)

V	Type	N [ɑ]_[ɑ]	N [i]_[i]	N [u]_[u]	N (flanker)
[æ]	L, F	6	6	6	0
[ɑ]	L, B	12	7	6	905
[ɑː]	L, B	9	6	6	0
[ɛ]	M, F	6	6	6	0
[ɛː]	M, F	3	3	3	0
[œ]	M, F	3	3	3	0
[e]	M, F	6	6	6	0
[eː]	M, F	6	6	6	0
[ɵ]	M, C	3	3	3	0
[ə]	M, C	6	6	6	0
[ʌ]	M, C	3	3	3	0
[ɔ]	M, B	6	6	6	0
[ɔː]	M, B	3	3	3	0
[o]	M, B	6	6	6	0
[oː]	M, B	6	6	6	0
[ɪ]	H, F	3	3	3	0
[i]	H, F	6	12	6	907
[iː]	H, F	6	9	6	0
[y]	H, F	3	3	3	0
[ɨ]	H, C	6	6	6	0
[ɨː]	H, C	6	6	6	0
[ʊ]	H, B	3	3	3	0
[u]	H, B	6	6	12	928
[uː]	H, B	6	6	9	0

Magnitudes of movement (MMs) for each of the two speakers were first subjected to a z-score normalization within speaker, to allow pooling of the magnitude data. Within a vowel token, it is likely that the MM from one frame is highly correlated with the MM from adjacent frames. To reduce this collinearity, the normalized magnitude scores for all the frames that corresponded to a single token were averaged, to produce a single average normalized magnitude of movement measure for each vowel token in the data.

3. RESULTS

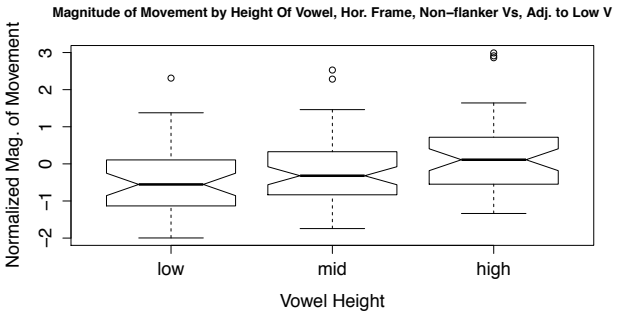
In order to interpret the MM values meaningfully, both the identity of each vowel and the context in which it was spoken must be taken into account. That is, we should not *a priori* expect the MM of the tongue to be the same in a high vowel as a low vowel, nor for the MM in a high vowel to be the same when it is adjacent to another high vowel as compared to when it is adjacent to a low vowel.

Therefore, we present two types of analysis for both height and backness: first, the average MMs in a variety of *target* vowels, when they are adjacent to a limited type of *flanker* vowel; second, the average MMs in the limited set of *flanker* vowels when they are adjacent to a variety of *target* vowels.

3.1. Correlates of vowel height

Figure 2 shows the average MMs of target vowels, divided by whether they themselves were high, mid, or low vowels, that were adjacent to the low flanker [ɑ]. An α of 0.05 is assumed for all statistical tests.

Figure 2: Magnitudes of movement for high, mid, and low target vowels, adjacent to [ɑ]

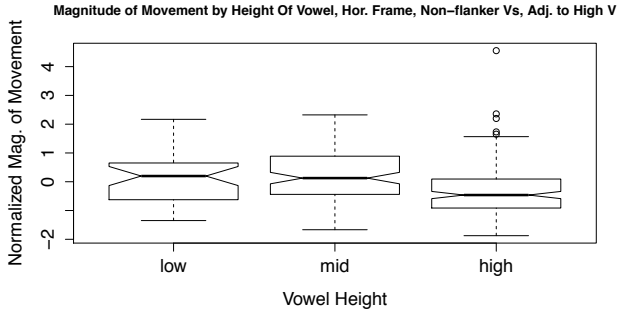


The average MM increases with vowel height, when the flanker is itself low. This is not surprising; one would expect that the production of a low vowel adjacent to another low vowel would involve relatively small tongue movements, while the production of a high vowel in the same context would involve greater tongue movement. Although this is the overall trend, and an ANOVA reveals that MM does significantly vary by height for these [ɑ]-adjacent vowels [$F(2,147)=4.78$, $p = 0.01$], the

difference between low and mid vowels is not statistically significant, as revealed by a planned comparison t-test [t(97)=1.38; n.s.]. The differences between low and high vowels [t(79)=2.84, $p < 0.01$; Cohen's d=0.65] and between mid and high vowels [t(118)=2.16, $p < 0.05$; Cohen's d=0.40] are each significant, with moderate effect sizes.

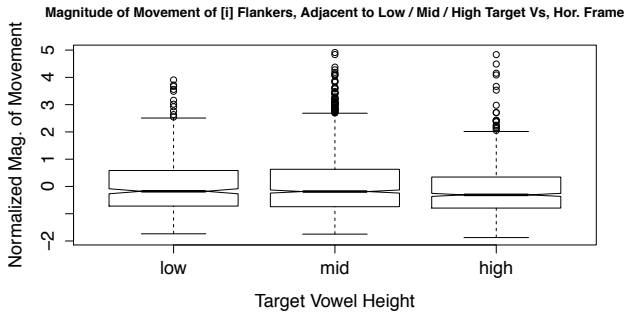
The lack of effect between low and mid vowels does not seem to be simply an effect of sparse data not confirming a trend; the average MMs in the same target vowels show an inverse pattern when they are adjacent to a high flanker vowel ([i] or [u]), as in Figure 3 [F(2, 376)=13.34 , $p<0.001$]. Thus, a *high* vowel produced adjacent to a high vowel shows the smallest MM, significantly lower than either low vowels [t(230)=3.36, $p < 0.001$; Cohen's d=0.57] or mid vowels [t(334)=4.759, $p < 0.001$; Cohen's d=0.52]. There is again no difference between the MMs for mid vs. low vowels adjacent to a high vowel [t(188)=0.10; n.s.].

Figure 3: Magnitudes of movement for low, mid, and high target Vs, adjacent to [i], [u]



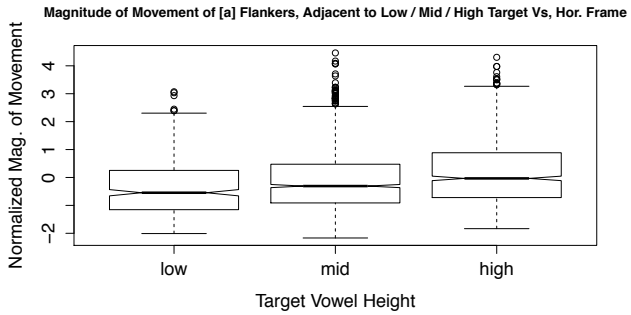
We also measured the magnitude of movement in a single flanker when it is adjacent to a variety of targets. Figure 4 shows the results for the flanker [i], and Figure 5 for the flanker [ɑ] (note that for the [ɑ] flanker, one outlying data point that had a magnitude of movement more than 7 standard deviations from the mean was removed).

Figure 4: Magnitudes of movement for [i], adjacent to low, mid, or high target Vs



The pattern for [i] matches the patterns seen in Figures 2 and 3; overall, height is a significant factor in predicting MM [F(2,3003)=13.4, $p < 0.001$], and the vowel [i] is produced with the smallest average MMs when adjacent to other high vowels, though the effects are small [vs. low V: t(902.63)=3.04, $p = 0.002$; Cohen's d=0.17; vs. mid V: t(2442.52)=5.29, $p < 0.001$; Cohen's d=0.21]. There is, once again, no difference in the MMs associated with the production of this high vowel when it is adjacent to a low vs. a mid vowel [t(941.12)=0.96; n.s.].

Figure 5: Magnitudes of movement for [ɑ], adjacent to low, mid, or high target Vs



The only place where a difference is seen between the low and mid vowels is when the MMs are measured within [ɑ], adjacent to vowels at each of the three heights, as shown in Figure 5. Again, there is a significant overall effect of height [F(2,2874)=40.7, $p < 0.001$]. In this case, the smallest MMs are for low vowels, which are significantly smaller than for mid vowels [t(666.20)=3.87, $p < 0.001$; Cohen's d=0.21], which in turn are significantly smaller than for high vowels [t(2483)=6.82, $p < 0.001$; Cohen's d=0.28].

Thus, while the correlation between MM and vowel height is in the anticipated direction, with increased MMs associated with larger height differences, there seems to be a larger or clearer distinction between high and non-high vowels than between low and non-low vowels.

3.2. Correlates of vowel backness

We next examine how MMs correlate with vowel backness, using the same two types of measures (targets adjacent to different flankers, and flankers adjacent to different targets). Figure 6 shows the average MMs of target vowels, divided by whether they themselves were front, central, or back vowels, that were adjacent to the front flanker vowel [i].

Again, there is an overall significant effect of backness [F(2, 148)=4.45, $p=0.01$], and the direction of results aligns with intuition, such that front targets produced adjacent to the front flanker [i] show the

smallest average MMs, significantly smaller than either the MMs for central vowels [t(88)=2.43, p = 0.02; Cohen's d=0.58] or back vowels [t(125)=2.65, p = 0.01; Cohen's d=0.47]. The central and back vowels do not differ significantly, however [t(83)=0.40; n.s.].

Figure 6: Magnitudes of movement for front, central, and back target Vs, adjacent to [i]

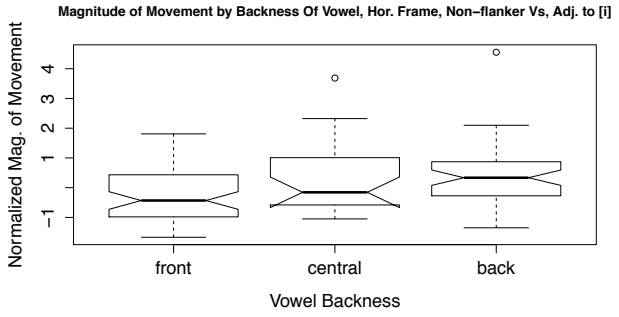
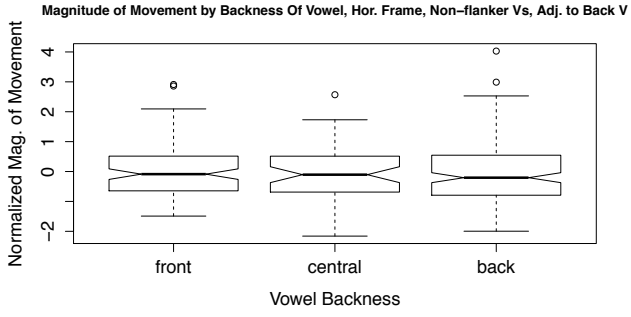


Figure 7 shows the average MMs of target vowels adjacent to the back flanker vowels, [ɑ] and [u]. The trend is generally as expected, though it is not statistically significant. An ANOVA indicates no significant effect of vowel backness on MMs for target vowels adjacent to back vowels [F(2,394)=0.01; n.s.]. This particular lack of difference may be due in part to the overall differences in degree of articulatory "backness" associated with a high back vowel like [u] vs. a low back vowel like [ɑ].

Figure 7: Magnitudes of movement for front, central, and back target Vs, adjacent to [ɑ], [u]



Figures 8 and 9 show the average MMs for the flanker vowels [i] and [ɑ], respectively, when they are adjacent to front, central, or back target vowels.

The back vowel [ɑ] shows the most movement when adjacent to front vowels and the least when adjacent to back vowels, while the front vowel [i] shows the inverse pattern. ANOVAs show overall effects for backness on MMs for each flanker vowel [[ɑ]: $F(2,2874) = 62.27$, $p < 0.001$; [i]: $F(2,3003) = 33.01$, $p < 0.001$]. All pairwise comparisons are statistically significant, with effect sizes ranging from small (Cohen's d = 0.15) for central/non-central comparisons to medium (Cohen's d = 0.46) for front/back comparisons; the effect sizes were slightly larger overall for the [ɑ] comparisons than the [i] comparisons.

Figure 8: Magnitudes of movement for [i], adjacent to front, central, or back target Vs

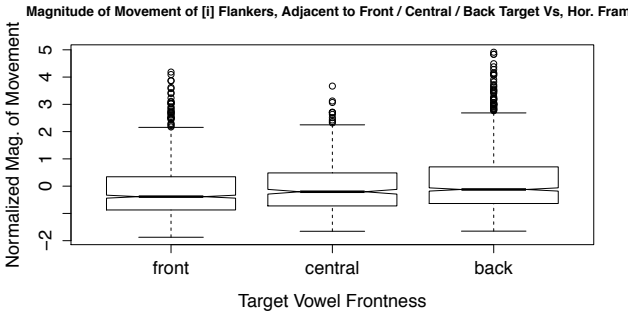
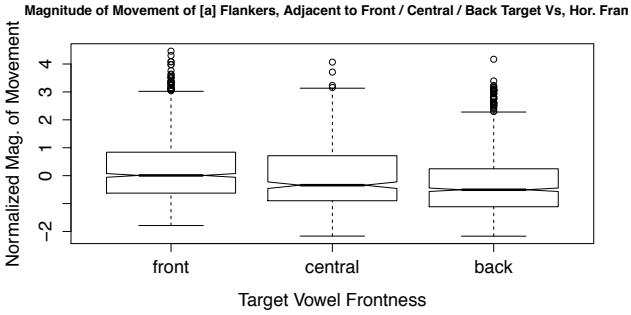


Figure 9: Magnitudes of movement for [ɑ], adjacent to front, central, or back target Vs



4. CONCLUSION

This paper has shown that the average magnitudes of movement of the tongue, from frame to frame in an ultrasound video, as measured by optical flow analysis, generally align with the intuitions one might have based on the phonological categories of height and backness. The specific magnitudes are highly dependent on the context in which the vowels are produced. For both height and backness, the clearest differences in MMs are found when looking at the productions of a single vowel quality across contexts that contain multiple vowel types, rather than aggregating the MMs of a variety of types in single contexts. Furthermore, height differences seem to be more robust to this distinction in measurement techniques than backness ones. Within vowel height, the clearest distinctions are between high and non-high vowels. A three-way distinction is more apparent in vowel backness measures, but only when examining individual vowel qualities.

5. REFERENCES

[1] Barbosa, A.V. p.c., 2013. FlowAnalyzer. [Computer Program].

[2] Barbosa, A.V. & Vatikiotis-Bateson, E. 2014. Optical flow analysis for measuring tongue-motion. Paper presented at the 168th Meeting of the Acoustical Society of America, Indianapolis, IN.

[3] Barbosa, A.V., Yehia, H.C., Vatikiotis-Bateson, E. 2008. Linguistically valid movement behavior measured non-invasively. In: Göcke, R., Lucey, P., Lucey, S. (eds), *Auditory and visual speech processing - AVSP08*. Moreton Island, Australia: Caual Productions, 173-77.

[4] Boersma, P., Weenink, D. 1992-2014. Praat: a system for doing phonetics by computer.

[5] Fleet, D.J., Weiss, Y. 2006. Optical Flow Estimation. *Handbook of Mathematical Models in Computer Vision*, ed. N. Paradgios, N., Chen, Y., & Faugeras, O.D. Secaucus, NJ: Springer, 239-57.

[6] Hall, K.C., Smith, H., McMullin, K., Allen, B., Yamane, N., & Gambarage, J. 2014. Articulatory correlates of phonological relationships. 14th Meeting of the Association for Laboratory Phonology.

[7] Horn, B.K.P., Schunck, B.G. 1981. Determining optical flow. *Artificial Intelligence* 17, 185-203.

[8] Mielke, J. 2012. A phonetically based metric of sound similarity. *Lingua* 122.145-63.

[9] Moisik, S., Lin, H., & Esling, J.H. 2014. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association* 44.21-58.

[1] We give special acknowledgment to Jeff Mielke for his help with providing the recordings for this paper.