DIFFERENCES IN DIPHTHONG PRODUCTION BETWEEN L1 AND L2 SPEAKERS OF HUL’Q’UMI’NUM

Sky Onosson[1] and Sonya Bird[2]

[1] University of Manitoba, [2] University of Victoria
Sky@Onosson.com, sbird@uvic.ca

ABSTRACT

This paper examines the acoustic properties of Hul’q’umi’num’ vowel-glide sequences [ej, ew] as well as short and long [e, eː], comparing pronunciations of a single L1 speaker to those of a group of fifteen L2 learners who are native speakers of English. Generalized Additive Models (GAMs), which permit statistical comparisons of non-linear data such as transitional formant trajectories, were used in this study to investigate dynamic changes in acoustic qualities over time. From our results, we identify three key areas within which Hul’q’umi’num’ learners differ significantly from the L1 speaker: vowel duration, vowel and glide articulatory target positions, and dynamics of the intensity contour. This documentation work lays the foundation for creating pedagogical resources focused on teaching and learning pronunciation, as part of ongoing, collaborative Hul’q’umi’num’ language revitalization efforts.

Keywords: Hul’q’umi’num; diphthongs; articulation; GAMs; language revitalization.

1. INTRODUCTION

Hul’q’umi’num’ territory extends along the Salish Sea from Nanoose to Malahat on Vancouver Island and neighbouring islands. Very few (approximately 40) first language speakers remain, but there are over 200 fluent second language speakers and over 1,000 leaners of all ages. Many of these leaners are currently at intermediate levels of proficiency and ready to tackle the more complex features of their language, including the details of pronunciation.

While there is much interest in teaching and learning ‘authentic’ pronunciation, resources are limited: popular pedagogical approaches in the Indigenous language revitalization context do not emphasize pronunciation; descriptions of pronunciation are rare and most often inaccessible to community members (written by and for linguists), and few opportunities exist for learners to interact with fluent speakers [2]. To support community-based pronunciation work, we are in the process of documenting the pronunciation features of first and second language speakers as well as working with elders, teachers, and learners to identify the perceived challenges for learners, and how best to overcome them.

One aspect of pronunciation that elders and teachers have noticed to differ between elders and learners involves the vowel-glide (VG) sequences /ej, ej’, ew, ew’/ (/j’, w’/ are glottalized resonants). Similar discrepancies in diphthong production across generations have been observed in other indigenous minority languages e.g. in Māori (New Zealand), which have been attributed to the influence of English [4; 9]. This study is aimed at delineating the particular areas where elders and learners differ in terms of VG production, with the goal of providing data for incorporation into pedagogical material to assist learners in developing more authentic pronunciations. Because the VG sequences under consideration share a common nucleus [e], we also investigated production of both short /e/ and long /eː/, which are phonologically distinct in Hul’q’umi’num’.

Note that, in this paper, we refer to [ej, ew] as VG sequences rather than diphthongs. We do this based on phonological properties not discussed in this paper. Phonetically, Hul’q’umi’num’ VG sequences may well be equivalent to what are considered diphthongs in other languages.

2. METHODS

2.1. Speakers

Speakers included a single first language speaking elder and a group of fifteen adult second language learners, ranging in age from early 20s to 60+. The elder is involved in all aspects of Hul’q’umi’num’ language documentation and revitalization, including supporting second language learners in their pronunciation work. The learners have a range of backgrounds with respect to language use, and varying levels of oral proficiency, from beginner to intermediate. Because the elder is female, only female learners were included in this study, so as avoid the necessity to normalize formant values across different-sex speakers.

2.2. Materials

Words containing [e, eː, ej, ew] were extracted from recordings of a larger (30-item) word list, designed to assess pronunciation challenges for Hul’q’umi’num’ learners. The word list did not contain any VG sequences with plain glides so, in this preliminary study, we made do with words containing /ej’, ew’/ sequences. Word-finally, Hul’q’umi’num’ glottalized resonants are generally pronounced as modal-voiced resonants followed by a full glottal stop; it was therefore easy to extract only the target [ej, ew] sequences for analysis (see below). In total four words, one per target sound/sequence, were included in the study, as shown in Table 1.

Table 1: Elicited words; /’/ indicates glottalized resonants.

Vowel	Word	
[e]	/ˈleləm’/	*house*
[eː]	/ˈʔeː’nθə/	*me*
[ej]	/sqʷəˈmej’/	*dog*
[ew]	/sqəˈl’ew’/	*beaver*

2.3. Procedure

The procedure for eliciting tokens was as follows: the elder sat with one learner at a time, in a quiet room, and went over the full 30-word list in sequence. For each word, the elder first read the word (reading task), and the learner repeated it (oral imitation task). Two repetitions per word were recorded in the following sequence: elder, first repetition > learner, first repetition > elder, second repetition > learner, second repetition. Recordings were made in Audacity [1], using a Yeti USB microphone in cardioid mode connected to an Apple iMac computer, and saved as 48 kHz, 16-bit uncompressed .wav files.

2.4. Acoustic analysis

Following the procedure outlined above, the elder’s dataset was compiled from that single speaker pronouncing each word twice in each of the 15 learner sessions (2 repetitions x 15 sessions x 4 words = 120 tokens); the learners’ dataset included 15 learners pronouncing each word twice in a single session (2 repetitions x 15 learners x 4 words = 120 tokens). This yielded a grand total of n=240 tokens which were included in the analysis.

Tokens of [e, eː, ej, ew] were manually segmented and transcribed in Praat [3]. Following segmentation, a Praat script [12] was utilized to extract acoustic measurements of total duration, and discrete formant measurements taken at 5% duration intervals throughout the vowel or VG sequence; the script was further modified to extract similar 5%-interval measurements of spectral intensity. This acoustic data was exported to R [7] for statistical testing and modelling. In order to compare curvilinear formant and integrity trajectories, generalized additive models or GAMs [5; 6] were implemented in R using the package itsadug [10].

3. RESULTS

The primary focus of this study concerns the VG sequences: [ej, ew]. However, we will present evidence that long [eː] has certain diphthongal characteristics, meriting its inclusion among these; results pertaining to short [e] are only discussed where especially relevant. We consider three areas of acoustic analysis in this section as follows: duration, formant trajectories, and intensity trajectories.

3.1 Duration

Mean durations and standard deviations for each vowel across elder (L1) and learners (L2) are summarized in Table 2.

Table 2: Mean vowel durations (ms) and standard deviations across speakers.

Vowel	L1 duration (s.d.)	L2 duration (s.d.)
[e]	160.8 (15.3)	163.2 (39.1)
[eː]	202.9 (28.4)	197.9 (45.4)
[ej]	177.3 (21)	153.5 (34.5)
[ew]	202.2 (28.5)	188.3 (28)

While learners produce broadly similar durations as the elder, they uniformly produce durations which are less extreme: compared to the elder, [ej, ew, eː] are shorter, while [e] is (slightly) longer. The duration of [ej] is noteworthy across speakers: for the elder, [ej] is shorter than both [eː] and [ew], only 17ms longer than [e]; for the learners, [ej] has the shortest duration of all, shorter even than monophthongal [e]. This leads to [ej] exhibiting the largest difference in mean duration between the elder and the learners (>20 ms) albeit within one standard deviation.

3.2. Formant trajectories

We conducted statistical comparisons of F1 and F2 formant trajectories in GAMs. For each token, 20 discrete per-formant measurements were taken at 5% intervals, allowing us to monitor variation between speaker groups across the entire trajectory of the vowel in high resolution. GAMs comparisons calculate “smooths” representing the mean formant trajectory, plotted as a solid line, accompanied by shaded regions representing confidence intervals associated with the distribution of formant values across the tokens comprising the dataset under consideration. Where the confidence intervals across two conditions—in this case, L2 language learners vs. L1 elder—do not overlap, this indicates a statistically significant difference at that position (where confidence intervals do overlap, there is less certainty about whether or not the two conditions are statistically different or not). The formant comparisons for the VG sequences [ej, ew] as well as long [eː] are presented in Figures 1 through 3 below; the short [e] GAMs comparison did not present compelling differences between groups and so are not presented.

Overall, articulatory targets in VG sequences are closer together for learners than for the elder, especially with respect to height (F1). This results in less steep transitions between the vowel and the glide targets for learners than for the elder.

Figure 1: GAMs comparisons: formants of [ej].

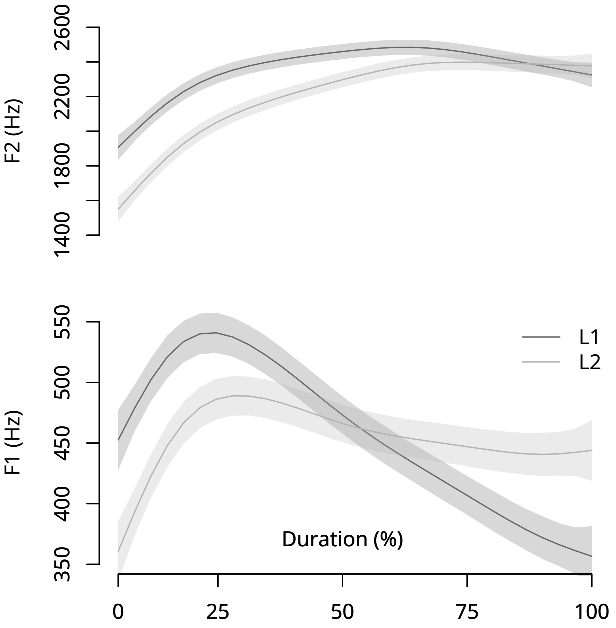


Figure 2: GAMs comparisons: formants of [ew].

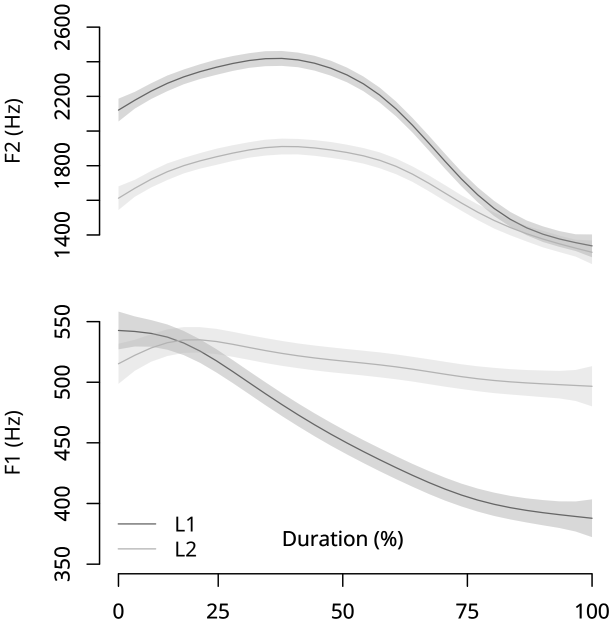
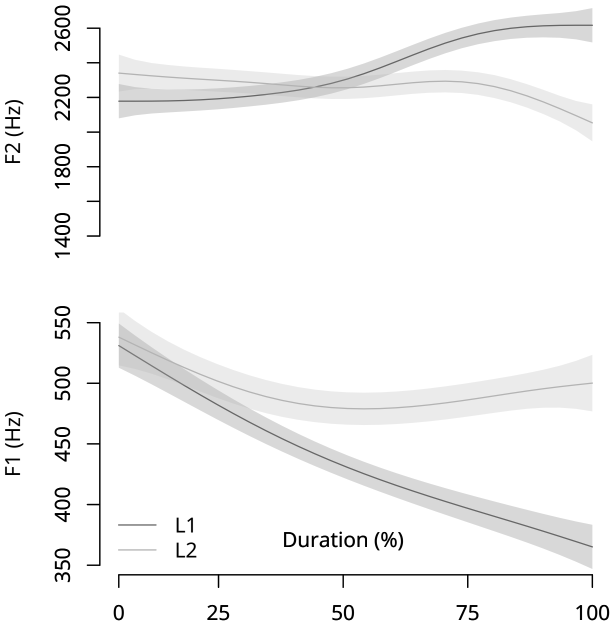


Figure 3: GAMs comparisons: formants of [eː].



In terms of height (F1), learners generally start at a similar height to the elder, but do not hit as high of a target (as low an F1) for the glide as the elder does. Thus, their glides [j] and [w] are more similar in height to that of nucleus [e], in comparison with the elder, hence less steep transitions.

In terms of backness (F2), the learners differ most substantially from the elder during the nucleus, being consistently further back (lower F2), but reaching the same F2 target for the glide as the elder by 100% of duration.

Interestingly, the elder's pronunciation of [eː] is relatively diphthongized (see Figures 1 and 3), across both formants, with two distinct targets. Viewing this vowel in parallel with the VG sequences, in terms of F1 the differences between the learners and the elder are similar with [eː] as they are for [ej] and [ew]; the learners' trajectories are relatively flat in terms of height, i.e., not as diphthongal as the elder's.

In terms of F2, the pattern for [eː] differs from the VG sequences in that the learners start at roughly the same nuclear position as the elder, and afterwards the two trajectories move in different directions; the elder's F2 rises (more front articulation), while the learners' lowers (retracts).

3.3. Intensity trajectories

GAMs comparisons of intensity trajectories were conducted following the method described for formant trajectories, with per-group intensity values normalized to the global mean. Again, there were no compelling differences between groups with regard to short [e].

Figure 4: GAMs comparison: intensity of [ej].

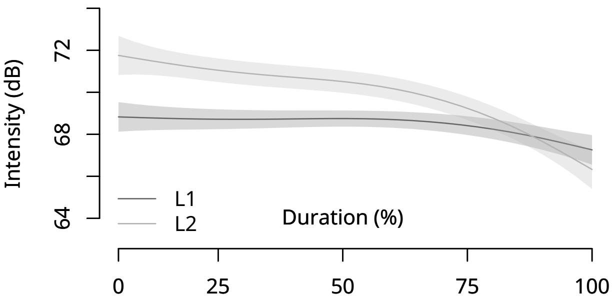


Figure 5: GAMs comparison: intensity of [ew].

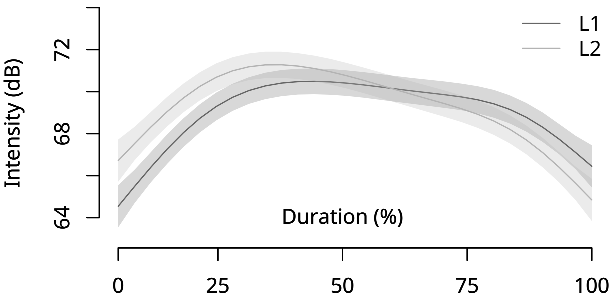
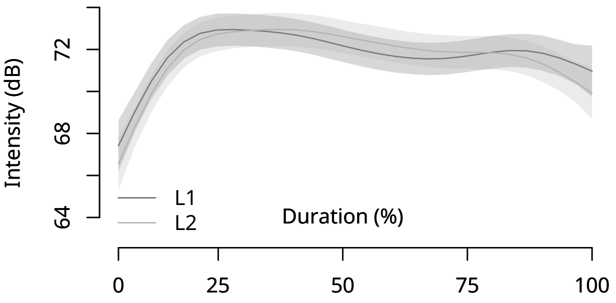


Figure 6: GAMs comparison: intensity of [eː].



The general trend is for intensity to drop off sooner for the learners than for the elder, across both VG sequences as well as (minimally) [eː]. Interestingly, both [ew] and [eː] seem to have two intensity peaks for the elder, one prior to 50% duration and one after 75%, suggesting two relatively distinct components of the sound (sequence), adding further evidence that [eː] is fairly diphthongal. In contrast, [ej] does not exhibit an obvious "two-peak" intensity curve for the elder. It is notable that the learners appear to be replicating the two-peak pattern fairly closely, for both [ew] and [eː], albeit with the earlier intensity drop-off previously mentioned.

4. DISCUSSION

Our analysis indicates that certain acoustic features of Hul'q'umi'num' VG sequences are quite similar between learners and the elder, while others are more distinct. In terms of duration, learners' relative vowel-to-vowel durations are similar to the elder's, but their mean per-vowel durations are briefer than the elder's, most substantially for [ej]. In terms of formant trajectories, learners' VG sequences are less transitional than the elder's: they are more retracted during the nucleus (F1) and raise less during the glide (F2). For diphthong-like [eː], in addition to producing a more stable vowel, learners also exhibit a mismatch with the elder in terms of the glide front-back position. Finally, in terms of acoustic intensity, learners exhibit a fairly close match with the elder's production, especially for the two vowels [ew] and [eː] which exhibit a "two-peak" intensity contour, although for all vowels their final intensity drop-off tends to occur slightly earlier than the elder.

In summary, our measurements show that, compared to the elder, learners produce VG sequences that tend to be less transitional, shorter, and with earlier drop-offs in intensity; in short, learner's productions are more reduced. While this pattern is relatively clear and consistent across acoustic parameters, the explanation is less certain. It could be that learners are hypo-articulating, perhaps under the influence of English. Conversely, it could be that the elder is hyper-articulating in this particular teaching-learning context [8, 9]. To assess these competing explanations, a more thorough understanding is needed of (a) VG sequences in the local variety of English (as a possible influence for learners in particular) and (b) Hul'q'umi'num' VG sequences in other, more naturalistic speech contexts.

From this preliminary study, it is not entirely clear how [ej] and [eː] are best described, independently and also in relation to one another. The VG sequence [ej] has a shorter duration and a simpler intensity curve than both [ew] and [eː], implying that it may not be a two-sound (VG) sequence in the same way [ew] is. Conversely, [eː] has a similar duration and intensity contour as [ew] and has similar a degree of formant transitionality as [ej], implying that it may actually be closer to a two-sound sequence than [ej]. Thinking about possible English influence (through second language learning and/or language contact), it is possible that the line is blurred in Hul'q'umi'num' between [ej] and [eː] because both correspond to monophthongal /e/ in English. More comprehensive research is needed to understand how /e, eː, ej, ej', ew, ew'/ should be characterized in relation to one another, including production and perception studies of these sounds/sequences in more controlled environments and across a broader range of speakers.

This preliminary study was driven by discussions with Hul'q'umi'num' elders and teachers about variation they have noticed in the pronunciation of VG sequences. The documentation work we have done here is a first step in understanding this variation, and providing the foundation for deciding whether and how to approach this variation—and variation more broadly—in teaching the language.

5. REFERENCES

[1] Audacity Team. 2018. Audacity(R): Free Audio Editor and Recorder (Version 2.3.0). https://www.audacityteam.org/

[2] Bird, S., Miyashita, M. In press. Teaching phonetics in the context of language revitalization. *Proceedings of the 2nd International Symposium on Applied Phonetics*.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer (Version 6.0.43). http://www.praat.org/

[4] Harlow, R., Keegan, P., King, J., Maclagan, M., Watson, C. 2009. The changing sound of the Māori language. In Stanford, J. N., Preston, D. R. (eds.), *Variation in Indigenous Minority Languages*. Amsterdam/Philadelphia: John Benjamins, 129–152.

[5] Hastie, T., Tibshirani, R. 1987. Generalized Additive Models: Some Applications. *Journal of the American Statistical Association, 82*(398), 371–386.

[6] Hastie, T.J., Tibshirani, R.J. 1990. *Generalized Additive Models*. New York: Chapman and Hall.

[7] R Core Team. 2018. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/

[8] Saito, K., van Poeteren, K. 2012. Pronunciation-specific adjustment strategies for intelligibility in L2 teacher talk: results and implications of a questionnaire study. *Language Awareness, 21*(4), 369–385.

[9] Uther, M., Knoll, M.A., Burnham, D. 2006. Do you speak E-N-G-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49*, 2–7.

[10] van Rij, J., Wieling, M., Baayen, R., van Rijn, H. 2016. itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs (Version 2.2).

[11] Watson, C.I., MacLagan, M.A., King, J., Harlow, R., Keegan, P.J. 2016. Sound change in Maori and the influence of New Zealand English. *JIPA, 46*(2), 185–218.

[12] Xu, Y. 2015. FormantPro.praat (Version 1.4). http://www.phon.ucl.ac.uk/home/yi/FormantPro/