DIFFERENCES IN VOWEL-GLIDE 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 Hul'q'umi'num' speaker to those of a group of fifteen L2 speakers. 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 the L1 speaker differs from the L2 speakers: 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 language revitalization efforts.

Keywords: Hul'q'umi'num'; vowel-glide sequence; diphthong; 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) L1 speakers remain, but there are over 200 fluent L2 speakers and over 1,000 learners of all ages. Many of these learners are currently at intermediate levels of proficiency and ready to tackle the more complex features of their language, including the details of pronunciation.

To support language work within the community, the Hul'q'umi'num' Language Academy runs both undergraduate and graduate programs through a partnership between the Hul'q'umi'num' Language and Culture Collective and Simon Fraser University, facilitated by Dr. Donna Gerdts. These programs include Phonetics courses, which provide an ideal setting to document pronunciation across a range of speakers [2] and explore pedagogical approaches to teaching and learning [4].

One aspect of pronunciation that Gerdts and the elders she works with have noticed to differ between L1 and L2 speakers 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 [7, 14]. This study is aimed at delineating the particular areas where L1 and L2 speakers 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 L1 speaker and a group of fifteen adult L2 speakers. The L1 speaker is an elder, linguist and teacher who is involved in all aspects of Hul'q'umi'num' language documentation and revitalization, including supporting learners in their pronunciation work. The L2 speakers are students in the Hul'q'umi'num' Language Academy's undergraduate program and are L1 speakers of English, ranging in age from early 20s to 60+. They have a range of backgrounds with respect to language use, and varying levels of oral proficiency, from beginner to intermediate. Because the L1 speaker is female, only female L2 speakers 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 (indicated by /'/) 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

Vowel	Word	
[e]	/leləm'/	*house*
[eː]	/ʔeːn'θə/	*me*
[ej]	/sqʷəmej'/	*dog*
[ew]	/sqəl'ew'/	*beaver*

2.3. Procedure

The procedure for eliciting tokens was as follows: the L1 speaker sat with one L2 speaker at a time, in a quiet room, and went over the full 30-word list in sequence. For each word, the L1 speaker first read the word, and the L2 speaker repeated it. Two repetitions per word were recorded in the following sequence: L1 first repetition > L2 first repetition > L1 second repetition > L2 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 L1 and L2 datasets were each comprised of 2 repetitions x 15 sessions x 4 words = 120 tokens, yielding a grand total of n=240 tokens for analysis. It should be noted that, because the L1 tokens all came from a single speaker, while the L2 tokens are aggregated from fifteen distinct speakers, the statistical significance of the results from this preliminary study should be interpreted with caution.

Tokens of [e, eː, ej, ew] were manually segmented and transcribed in Praat [6]. Following segmentation, a Praat script [15] 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, which was included on the basis of impressions among elders and teachers that the relative "weight" (or prominence) of the vowel and glide was different in L1 and L2 speakers. Acoustic data was exported to R [10] for statistical testing and modelling. In order to compare curvilinear formant and intensity trajectories, generalized additive models GAMs [8, 9] were implemented in R with the package itsadug [13].

3. RESULTS

The primary focus of this study concerns the VG sequences [ej, ew]. Evidence that long [eː] has certain diphthongal characteristics merit its additional inclusion among these sequences. Results pertaining to short [e] are less similar and only discussed where especially relevant.

3.1 Duration

A two-way ANOVA test of duration by vowel type and speaker group (L1 vs. L2) revealed a significant effect of vowel ($p < 0.001$), but no significant effect for speaker group. Nonetheless, L2 speakers tended to produce less extreme durations than L1 speakers: in comparison to the L1 speaker, L2 speakers had shorter [ej, ew, eː] and a slightly longer [e]. The duration of [ej] is noteworthy across speakers: L1 [ej] is shorter than both [eː] and [ew], only 17ms longer than [e]; L2 [ej] has the shortest duration of all, shorter even than monophthongal [e]. This leads to [ej] exhibiting the largest difference in mean duration between L1 and L2 (>20 ms) albeit within one standard deviation.

Table 2: Mean vowel durations (ms)

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)

3.2. Formant trajectories

We conducted statistical comparisons of F1 and F2 formant trajectories in GAMs. As is standard in GAMs analysis, we used time-normalized durations. 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 for the two conditions—in this case, L1 vs. L2—do not overlap, this indicates a statistically significant difference at that position. 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 L2 vs. L1, especially with respect to height (F1). This results in less steep transitions between the vowel and the glide targets for L2. In terms of height, F1 generally starts at a similar level for L2 as L1, but does not descend as far for the glide. Thus, L2 glides [j] and [w] are more similar in height to that of nucleus [e], in comparison with L1. In terms of backness, F2 of L2 differs most substantially from L1 during the nucleus, being consistently lower (retracted articulation), but reaching the same F2 target for the glide as L1 by 100% of duration, for both vowel-glide sequences.

Interestingly, the L1 pronunciation of [eː] is relatively diphthongized (see Figure 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 L1 and L2 are similar for [eː] as they are for [ej] and [ew]; L2 trajectories are relatively flat in terms of height, i.e., not as diphthongal as L1. In terms of F2, the pattern for [eː] differs from the VG sequences in that F2 of L2 starts at roughly the same nuclear position as L1, and afterwards the two trajectories move in different directions; F2 of L1 rises (more front articulation), while for L2 it lowers (retracts).

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

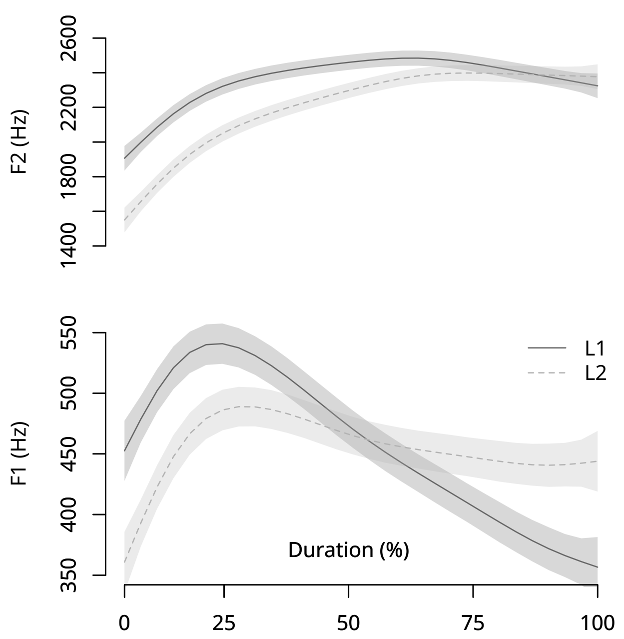


Figure 2: GAMs comparisons: formants of [ew].

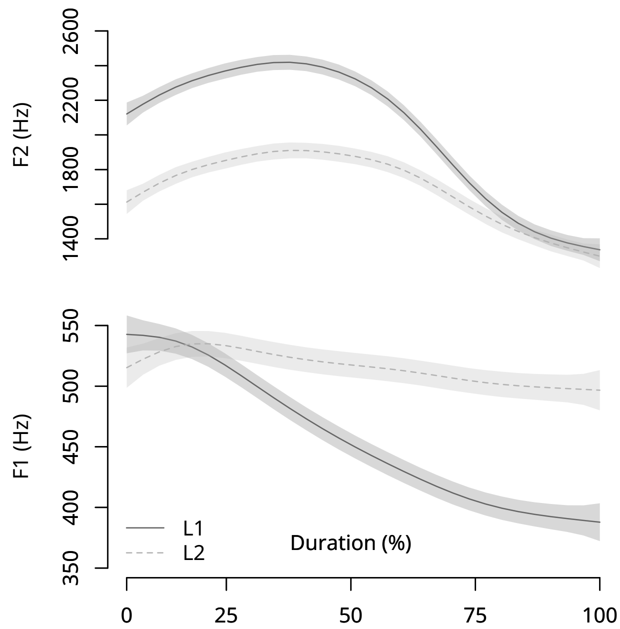
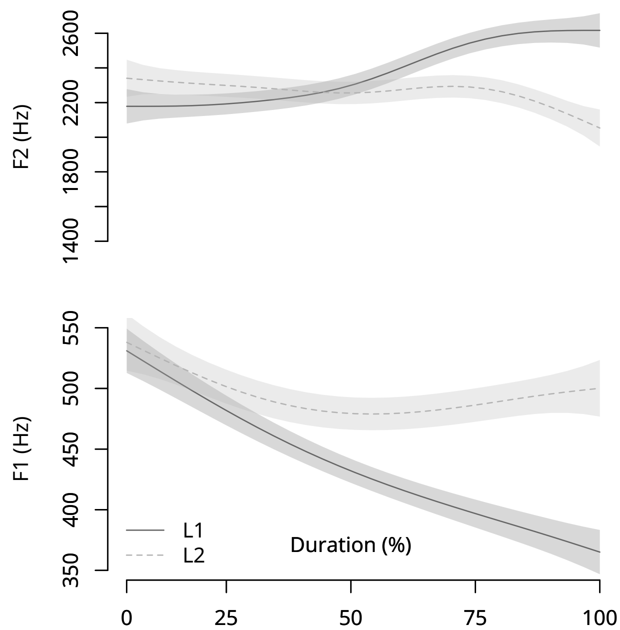


Figure 3: GAMs comparisons: formants of [eː].



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].

The general trend is for intensity to drop off sooner for L2 vs. L1, across both VG sequences as well as (minimally) [eː]. Interestingly, both [ew] and [eː] seem to have two intensity peaks for L1, one prior to 50% duration and one after 75%, suggesting two relatively distinct components of the sound (sequence). In contrast, [ej] does not exhibit an obvious “two-peak” intensity curve for the L1 speaker. It is notable that the L2 speakers appear to be replicating the two-peak pattern fairly closely, for both [ew] and [eː], albeit with the earlier intensity drop-off previously mentioned.

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

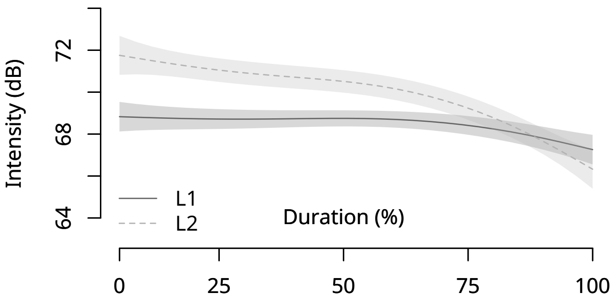


Figure 5: GAMs comparison: intensity of [ew].

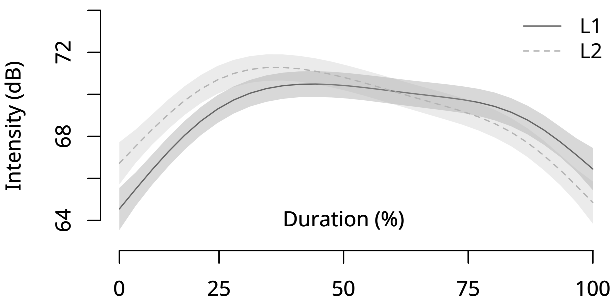
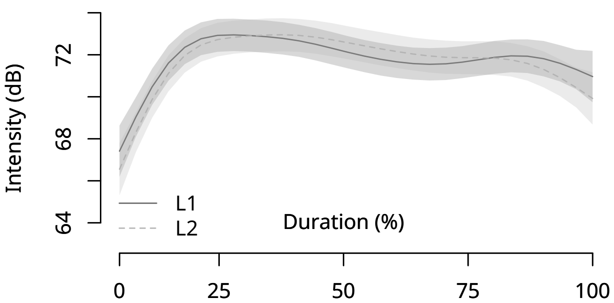


Figure 6: GAMs comparison: intensity of [eː].



4. DISCUSSION

Our analysis indicates that certain acoustic features of Hul’q’umi’num’ VG sequences are quite similar between L1 and L2 speakers, while others are more distinct. In terms of duration, L2 relative vowel-to-vowel durations are similar to L1, but their mean per-vowel durations are briefer than L1, most substantially for [ej]. In terms of formant trajectories, L2 VG sequences are less transitional than L1: 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, L2 speakers also exhibit a mismatch with L1 in terms of the glide front-back position. Finally, in terms of acoustic intensity, L2 speakers exhibit a fairly close match with L1 production, especially for [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 L1.

In summary, our measurements show that, compared to the L1 speaker, L2 speakers produce VG sequences that tend to be less transitional, shorter, and with earlier drop-offs in intensity; in short, L2 productions are more reduced. While this pattern is relatively clear and consistent across acoustic parameters, the explanation is less certain. It could be that L2 speakers are hypo-articulating, perhaps under the influence of English. Conversely, it could be that the L1 speaker is hyper-articulating in this particular task and/or teaching context [11, 12]. 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 L2 speakers 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 Hul’q’umi’num’ [ej] and [eː] are best described, independently and in relation to one another. [ej] has a shorter duration and a simpler intensity curve than both [ew] and [eː], implying that it may not be a two-sound sequence in the same way as [ew]. Conversely, [eː] has a similar duration and intensity contour as [ew] and a similar a degree of formant transitionality as [ej], implying that it may be closer to a two-sound sequence than [ej]. Thinking about possible English influence (through L2 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 English /e/. 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.

The documentation work we have done here is a first step in understanding one particular instantiation of phonetic variation, and providing the foundation for deciding whether and how to approach this variation—and variation more broadly—in teaching the language. Other aspects of pronunciation we are currently exploring include single segments (e.g. [k^{w}, k^{w}’, q^{w}, q^{w}’]), clusters (e.g. [st^{θ}’k^{w}] in /st^{θ}’k^{w}əʔiws/ ‘left hand’), and word-level and sentence level stress/intonation. The ultimate goal is to combine our expertise as phoneticians, Hul’q’umi’num’ linguists, and teachers, to create resources that support learners to meet their goal of achieving authentic Hul’q’umi’num’ pronunciation [3, 5].

5. REFERENCES

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

[2] Bird, S., Gerdts, D., Leonard, J. 2016. The realization of obstruents across speakers of Hul’q’umi’num’. *Canadian Acoustics, 44*(4), 134-135.

[3] Bird, S., Kell, S. 2017. The role of pronunciation in SENĆOŦEN language revitalization. *Canadian Modern Language Review 73* (4), 538-569.

[4] Bird, S., Luschiimtunaat Charlie, S., Claxton, R.A., S-hwutstus George, H., Percival, M., Sq’utxulenuhw Seymour, G. 2019. Seeing speech: Teaching and learning Hul’q’umi’num’ pronunciation with Praat. *ICLDC 6* Honolulu.

[5] Bird, S., Miyashita, M. 2019. Teaching phonetics in the context of language revitalization. *Proc. of the 2nd International Symposium on Applied Phonetics*, 39–44.

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

[7] 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.

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

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

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

[11] 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.

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

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

[14] 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.

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

* This project and the larger research program which it is part of would not be possible without the enthusiastic engagement of the Hul’q’umi’num’ Language Academy students, the Hul’q’umi’num’ Elders, our hosts at the Hul’q’umi’num’ Language and Culture Collective, and Dr. Donna Gerdts. Huy tseep q’a sii’em’!