BRITISH ENGLISH [KW], [K], AND [W] DISTINCTION IN BACK ROUND VOWEL CONTEXTS

Helena Beeley

University of Oxford
helena.beeley@phon.ox.ac.uk

ABSTRACT

Corpus data from the Audio British National Corpus was used to compare the acoustics of British English stops with velar and/or labial articulation in onset positions preceding back rounded vowels. The acoustics of labialized velar stops in British English have not commonly been described due to their frequent phonological analysis as clusters. Labialized velars have been characterized cross-linguistically in terms of low second formant frequency, but it was found that third formant frequency is a more significant cue to differentiation in back rounded vowel contexts. Plain velars with non-phonemic labialization resulting from proximity to a following back rounded vowel did not differ significantly from labialized velars in formant transition frequencies. Lower F3 in labial-velar approximants indicates more lip-rounding but higher F2 suggests a more fronted tongue position. Locus equations show that rounded vowels following [kw] are less affected by coarticulatory rounding than following [w].

Keywords: Phonetics; secondary labialization; labialized velars; corpus data.

1. INTRODUCTION

Due to the common analysis of English labialized velar stops as clusters rather than phonologically simple stops, acoustic analyses of these sounds are scarce and discussion of labialized velar sounds is often focussed on the labial-velar approximant [4, 6].

The primary acoustic effect of lip rounding has been described as lowering of the first three formant frequencies of the following vowel [4] or lowering of F1 and F2 [5]. The defining acoustic characteristic of labialized velars as described in [6] is the extreme lowering of F2 during release and the following vowel due to the locations of airstream velocity maxima in the vocal tract. Secondary labialization of stops is characterized in Tlingit, Spanish and Korean by low burst energy [7, 3]. However, the effect of labialization on rounded vowel transitions was not discussed in these studies, nor were English stops with secondary labialization the focus of study. This paper aims to analyse the acoustics of phonemic secondary labialization of stops and approximants and of non-phonemic coarticulatory labialization of stops in back vowel contexts. Due to the absence of a contrast between labialized velar stops and /kw/ clusters in syllable onsets in English the question of the exact phonological status of the sounds is not considered very relevant to the present discussion.

2. THE CORPUS

The Audio British National Corpus [2] consists of approximately 7 million words of spontaneous English speech. Recordings were collected during the period 1991 to 1994 using analogue audio casette tapes. The tapes were digitized in 2009 to 2010 and stored in .wav format and downsampled to 16 kHz. Transcriptions, originally produced by hand, were aligned with the audio using a forced aligner based on HTK. The corpus is searchable by Arpabet phone and by word, with each entry including the tape reference and start and end times within that tape for each word or phone. The alignments are published in Praat TextGrid format and in text list format.

2.1. Gender imbalance

Although initial recruitment resulted in a 1:1 ratio of female to male participants, in this research twice as many tokens from male speakers were found than tokens from female speakers. The effect on this paper was that, due to the relative lack of female tokens, analysis was limited to male speech.

3. METHOD

3.1. Token selection

Tokens were extracted for analysis by searching for instances of the required search terms, i.e. word-initial instances of ‘qu’ (Table 1). The start

Table 1: The top ten attestations of English words beginning with /kw/ in the Audio British National Corpus (BNC), before token selection.

Token	*n*
Quite	1947
Question(-s, -ed)	994
Quick(-ly, -er)	393
Quality	294
Quarter(-s, -ly)	243
Quid	132
Quarr(-y,-ies)	136
Quote(-s, -d)	157
Queen	54
Quiet	53

and end times provided by the forced alignment program were used to play each selection with an additional 0.1 s at each end to allow the researcher to make a judgement on 1) whether or not the selection contained appropriate audio, including filtering out words beginning with the letter sequence ‘qu’ that did not contain labiovelars, such as *quiche*; 2) quality of recording; 3) accuracy of alignment; and 4) sex of the speaker. Well-aligned selections were given a tag specifying male or female, excised, and stored as individual .wav files. Poorly aligned or bad quality tokens were discarded. The search program yielded approximately 8000 potential tokens, of which 4808 were of sufficient quality. For comparison, tokens with a phonemically equivalent vowel were selected with the target onset segments, although frequency was lower than for the [kw] tokens. These were: *port-*, including *porter*, *portable* etc., $n = 33$; *caught*, $n = 34$; *ward* and *wart*, $n = 30$; and *quart()*, $n = 180$.

3.2. Extraction of speech data

The vocalic portions of the .wav files were isolated using ESPS function *get_f0* [1] to calculate the probability of voicing (expressed as 0 or 1) at intervals of 5 ms throughout the clip. The voicing probability scores produced by *get_f0* were used to extract the first voiced section in each .wav token, equivalent to the first vocoid of the word; clips were tested at random to ensure they were sufficiently well-aligned for this method to be successful. Formant values from those timepoints were then extracted using the ESPS *formant* function. Five formants were extracted using an LPC order of 16 and a pre-emphasis constant of 0.9, due to the low signal-to-noise ratio of many tokens.

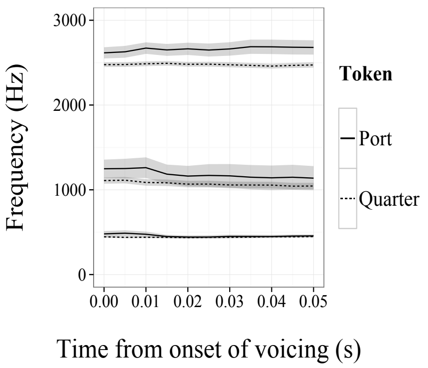
3.3. Locus equations

Locus equations were derived by plotting the F2 frequencies at voice onset against the F2 frequencies at the vowel target for each stop, determined here to be after 0.05 s of voicing (by which time it is expected the effects of stop coarticulation on the vowel are most reduced) and calculating a regression line [9, 8] using the *lm()* function in R statistics software. Locus equations have been used to acoustically categorise onset stops and represent the degree of coarticulation within the CV articulation; here they will enable comparison of degrees of coarticulation between different onset stops in the same vocalic environment. Locus equations have only occasionally been derived for labialized stops [3]; in this instance one would expect locus equations to indicate that labialized velar stops and labial-velar approximants are heavily coarticulated towards the following rounded vowel, because the stops are already rounded due to their secondary labial articulation.

4. RESULTS

4.1. Results: Formant transitions

Figure 1: Mean first to third formant transitions for the back round vowel in BNC tokens *port* ($n = 33$) and *quarter()* ($n = 180$) with 95% confidence region.



Mean formant frequencies were calculated to compare formant transition frequencies at different stages of vocoid articulation for each onset; *p* values were derived using Welch two sample *t* tests 2. F1, F2, and F3 are low in all the labialized stop contexts (Table 2) relative to the plain labial onset, but differences in F1 and F2 were not significant between any of the stops tested, indicating that

Table 2: Mean formant frequencies and p values for each onset in comparison to /kw/ at different stages of vocoid articulation. Significant p values are asterisked.

At the onset of voicing			
	Quarter()	*War(t,d)*	*p* value
F1 (Hz)	445	425	0.35
F2 (Hz)	1109	1267	0.07
F3 (Hz)	2476	2158	*0.001
At 0.05 s after the onset of voicing			
	Quarter()	*War(t,d)*	*p* value
F1 (Hz)	448	438	0.58
F2 (Hz)	1045	1192	0.10
F3 (Hz)	2473	2171	*0.001
At the onset of voicing			
	Quarter()	*Caught*	*p* value
F1 (Hz)	445	456	0.64
F2 (Hz)	1109	1055	0.46
F3 (Hz)	2476	2344	0.05
At 0.05 s after the onset of voicing			
	Quarter()	*Caught*	*p* value
F1 (Hz)	448	469	0.27
F2 (Hz)	1045	970	0.38
F3 (Hz)	2473	2434	0.54
At the onset of voicing			
	Quarter()	*Port(-)*	*p* value
F1 (Hz)	445	480	0.34
F2 (Hz)	1109	1248	0.23
F3 (Hz)	2476	2616	0.06
At 0.05 s after the onset of voicing			
	Quarter()	*Port(-)*	*p* value
F1 (Hz)	448	457	0.65
F2 (Hz)	1045	1137	0.53
F3 (Hz)	2473	2679	*0.03

Figure 2: Mean first to third formant transitions for the back round vowel in BNC tokens *quarter()* ($n = 180$) and *caught* ($n = 34$) with 95% confidence region.

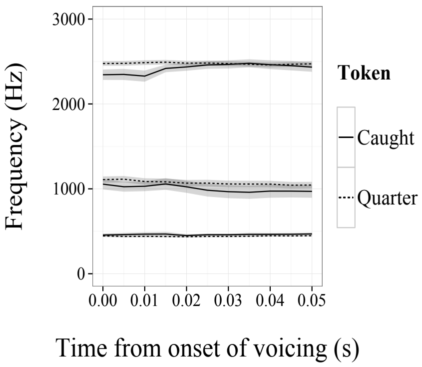
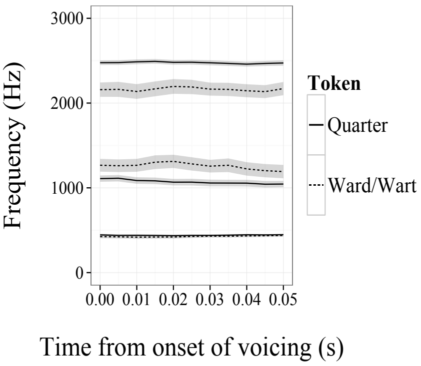


Figure 3: Mean first to third formant transitions for the back round vowel in BNC tokens *quarter()* ($n = 180$) and *ward* or *wart* (combined $n = 30$) with 95% confidence region.



F2 is not a prominent cue to distinction between transitions from labial stops, stops with secondary labialization, and stops with non-phonemic coarticulatory labialization. F3 remains significantly lower after the [kw] burst than after [p] for the duration of the vowel ([kw] $\mu = 2473$ Hz, [p] $\mu = 2679$ Hz, $p = 0.03$). F3 is also significantly lower after the labialized velar stop than after the labial-velar approximant (p = 0.01) throughout the duration of the vocoid while there is no difference in F2, indicating that F3 may be an important cue to perceptual distinction. Words in the *war()* category have the lowest frequency F3 transitions ($\mu = 2171$ Hz) indicating a greater degree of lip-rounding.

Although *caught* tokens had a slightly lower F3 frequency overall than *quarter()* tokens—which appears as a distinctive, sharply rising transition in Figure 2—formant trajectories did not differ significantly for the *quarter()* and *caught* tokens in F1 or F2 at any point during the vowel transition and differed only marginally significantly at onset for F3 (p = 0.05), indicating that vowel quality is not affected differently by secondary labialization and by coarticulatory labialization and so does not offer any additional perceptual cues to distinction between secondary and non-phonemic labialization.

4.2. Results: Locus equations

4.3. Discussion

Results agreed with those predicted in Section 3.3, with both [kw] and [w] having steep slopes corresponding to a high degree of coarticulation. However, the relationship between F2 during onset and target does differ between [kw] and [w] before a rounded vowel (Figures 4.2 and 4.2). Words with [w] have the highest intercept indicating the smallest degree of coarticulation, whilst words with [k] and [kw] are roughly equivalent, suggesting that

Figure 4: Locus equation for F2 of *caught* tokens with 95% confidence region.

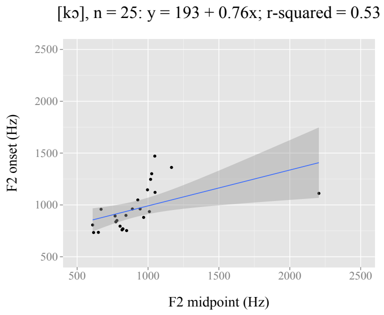


Figure 5: Locus equation for F2 of *quarter()* tokens with 95% confidence region.

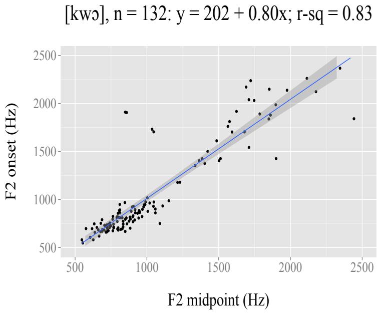
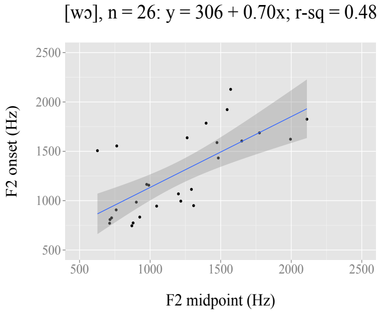


Figure 6: Locus equation for F2 of *war()* tokens with 95% confidence region.



the phonemic secondary labialization of [kw] does not lead to decreased coarticulation equivalent to that of [w] before a rounded vowel. Coarticulation towards a rounded vowel is less necessary for [w] indicating that roundedness equivalent to that of the rounded vowel is already present at the onset of voicing, unlike in the case of [kw] and [k].

The similar intercept values in locus equations for [kw] and [k] before back rounded vowels indicate a similar degree of coarticulation, which suggests that labialization is less pronounced for [kw] than for [w]; the labialized velar stop [kw] has more in common with the non-phonemically rounded [k] than it does with the phonemically rounded approximant [w].

Lip rounding is characterized by a lowering of F3, which is inhibited in the bilabial stop. The exceptionally low F3 of [w] transitions is indicative of a greater degree of labialization of [w] relative to [kw] and [k] as the oral cavity lengthens with lip rounding; however, F2 is not similarly lowered. High F2 is generally caused by tongue fronting: in this instance a possible explanation is that the relatively low F2 of [kw] transitions is assisted by the retraction of the tongue necessary for the velar articulation, whereas during the articulation of [w] the tongue can assume a position further forward in the vocal cavity.

Formant transitions do not provide strong acoustic cues to identification of [kw] and [k] before a back rounded vowel; velars with phonemic secondary labialization do not have a stronger coarticulatory effect on a following back rounded vowel than do velars without phonemic secondary labialization. Therefore it is likely that cues to distinguishing between the two are stronger in the burst phase of the stop. Acoustic similarity in formant transition frequency is likely to be a contributing factor in commonly attested historical delabialization of labialized velars in back round vowel contexts.

5. REFERENCES

[1] Entropic Signal Processing System (ESPS). *Entropic Research Laboratories, Inc.*

[2] Coleman, J., Baghai-Ravary, L., Pybus, J., Grau, S. 2012. Audio BNC: the audio edition of the Spoken British National Corpus. *Phonetics Laboratory, University of Oxford.*

[3] Denzer-King, R. 2013. The acoustics of uvulars in Tlingit. Master's thesis Rutgers University, New Brunswick.

[4] Ladefoged, P. 2001. *Vowels and Consonants*. Oxford: Blackwell.

[5] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

[6] Ohala, J., Lorentz, J. 1977. The story of [w]: an exercise in the phonetic explanation for sound patterns. *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society* volume 3.

[7] Suh, Y. 2008. The effect of secondary labialization on stop burst spectra. *The Journal of the Acoustical Society of America* 124(4), 2556–2556.

[8] Sussman, H. M., McCaffrey, H. A., Matthews, S. A. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*

90(3), 1309–1325.

[9] Sussman, H. M., Shore, J. 1996. Locus equations as phonetic descriptors of consonantal place of articulation. *Perception & Psychophysics* 58(6), 936–946.