Pre-aspiration in Bethesda Welsh: A Sociophonetic Analysis

Jonathan Morris
School of Welsh, Cardiff University
morrisj17@cardiff.ac.uk

Míša Hejná
School of Communication and Culture, Aarhus University
misa.hejna@cc.au.dk

Previous research has shown that pre-aspiration can be either a phonemic or variable linguistic feature susceptible to linguistic and extra-linguistic influences. In the case of Welsh, previous exploratory work has found the presence of pre-aspiration (Ball 1984; Morris 2010; Iosad Forthcoming; Spooner 2016), but the phonetic and phonological properties of this feature and its sociophonetic patterning in the language are not known. This paper presents analyses of the variety of Welsh spoken in Bethesda (Gwynedd). It reports the frequency of occurrence of pre-aspiration, its duration, and noisiness. As well as describing pre-aspiration, it attempts to ascertain the extent to which this feature is influenced by linguistic and extra-linguistic factors. Wordlist data were analysed from 16 Welsh–English bilinguals from Bethesda (Gwynedd, north Wales). Speakers were aged between 16 and 18 years old and the sample was stratified by speaker sex and home language (either Welsh or English). The results indicate that pre-aspiration is frequent in both fortis and lenis plosives (the latter of which are typically devoiced in Welsh). In addition to a number of linguistic influences on its production, both speaker sex and home language were found to be significant predictors of variation for some measures. The results are discussed with reference to previous studies of pre-aspiration in other languages and work on phonetic variation in Welsh-English bilingual speech.

1 Introduction

Pre-aspiration is broadly defined as a period of friction between the offset of modal voicing and the closure of a following voiceless obstruent (e.g. Laver 1994: 150; Helgason 2002: 9; Silverman 2003: 575; Gordeeva & Scobbie 2010). Although it is said to be a phonemic feature that cues a phonological contrast and precedes voiceless plosives in a relatively small number of languages (e.g. Icelandic, see Nance & Stuart-Smith 2013 for an overview), pre-aspiration is also found as a variable feature in a number of other languages and can be influenced by linguistic and extra-linguistic factors. In particular, recent work has shown that it is a feature of many varieties of English in Scotland (Gordeeva & Scobbie 2010; Clayton 2017), England (Jones & Llamas 2003; Watson 2007; Hejná & Scanlon 2015; Kettig 2015), Wales (Hejná 2015), and North America (Clayards & Knowles 2015).

In the case of Welsh, previous work has suggested that pre-aspiration is a non-obligatory feature of the language before fortis plosives (Ball 1984; Morris 2010; Iosad Forthcoming; Spooner 2016) but there has been no work which systematically examines this feature. The aim of this small-scale study, therefore, is to investigate pre-aspiration in the Welsh spoken in Bethesda (Gwynedd, north Wales) and the extent to which it is influenced by linguistic and extra-linguistic factors. Specifically, we present an analysis of fortis and lenis tokens (the latter of which are variably voiceless in Welsh, see Ball 1984: 18) and investigate (1) the frequency with which pre-aspiration occurs in the data, (2) the duration of pre-aspiration and (3) the noisiness of pre-aspiration using the band-pass filtered zero crossing rate measurement (BP ZCR; Gordeeva & Scobbie 2010; Nance & Stuart-Smith 2013).

1.1 Defining pre-aspiration preceding plosives

The friction which characterises pre-aspiration (Helgason 2002: 9) has been described as a turbulent flow of air passing through the glottis and occurs as a result of glottal abduction (following the end of modal voicing of the vowel) prior to closure of the following voiceless plosive (Ladefoged & Maddieson 1996). The abduction of the glottis results in a period of breathy voice prior to the offset of voicing and period of voiceless pre-aspiration. Breathy voice and pre-aspiration have been shown to be sensitive to different prosodic and segmental factors and might therefore be treated as separate phenomena (Kingston 1990; Hejná 2015). Breathiness has, however, also been shown to be an important cue in the perception of pre-aspiration (Ní Chasaide 1985). The majority of studies therefore adopt a broader definition of pre-aspiration which includes both the period of breathy voice at the offset of modal voicing and the following period of voiceless pre-aspiration (Helgason 2003; Helgason & Ringen

2008; Karlsson & Svantesson 2011; Nance & Stuart-Smith 2013; Ringen & van Dommelen 2013; Stevens 2010, 2011).

1.2 Fortis and lenis plosives in Welsh

Welsh has both fortis (/p, t, k/), and lenis (/b, d, g/) consonants which may appear word-initially, word-medially, or word-finally. The fortis series are typically voiceless and aspirated whereas the lenis series are thought to be variably voiced and unaspirated (Hannahs 2013: 15). Jones (1984: 41) notes that '[the lenis series] are regularly unvoiced in word initial and final positions […] and frequently so in a medial fully voiced environment'.

Devoicing of lenis plosives is found as a regular phonological process in Welsh. This is known as provection. Hannahs (2013: 150–154) distinguishes between two types of provection. The first type occurs at morpheme boundaries and in the formation of compounding and is reflected orthographically (e.g. *gwlyb* 'wet' [gwlɨːb] > *gwlypaf* 'wettest' [gwləppav]). The second type of provection affects lenis plosives in certain south-eastern varieties of Welsh and is affected by word position and stress (this is known as *calediad* in Welsh, see Thomas 1988). In such varieties, voiced obstruents are said to become voiceless following a stressed vowel and preceding another vowel or sonorant (e.g. *cegin* 'kitchen' [keˑkɪn], see Awbery 1984: 69; Hannahs 2013: 151).

In addition to post-aspiration, fortis consonants in word-medial position are also considered to be geminated. This occurs when voiced obstruents lengthen following a penultimate stressed vowel (e.g. *hapus* 'happy' [happɨs], see Ball & Williams 2001: 25; Hannahs 2013: 21). In northern varieties of Welsh (unlike in southern varieties), the penultimate stressed vowel is invariably short preceding both fortis and lenis consonants (Awbery 1984: 75). It is not known, however, whether there are durational differences between the duration of fortis plosives and voiceless lenis plosives in word-medial positions. In the context of the current research, it is important to ascertain whether there are differences between the duration of the voiceless fortis and lenis consonants in order to ensure that any differences found in the production of pre-aspiration between the two series are not linked to consonant duration. We therefore examine this further in Section 3.1.

There are few acoustic studies of voicing in other varieties of Welsh and, in particular, northern varieties, where the second type of provection referred to above does not occur. Consequently, it remains to be seen to what extent lenis plosives are voiced, whether voiceless lenis plosives in word-medial position are geminates, and to what extent pre-aspiration occurs in voiceless lenis plosives. We consider these questions in this paper in order to provide a more detailed account of pre-aspiration in Bethesda Welsh.

1.3 Pre-aspiration in Welsh

Ball (1984: 18) first suggested the presence of pre-aspiration in Welsh in an analysis of Voice Onset Time. Morris (2010) examined the frequency and duration of pre-aspiration before /p/, /t/, and /k/ in the speech of four Welsh–English bilinguals from different areas of north Wales. In this preliminary study, based on 119 Welsh tokens and 49 English tokens, he found that 60.5% (*n*=72) of the Welsh tokens were pre-aspirated. Similar to the results of previous studies across languages (e.g. Foulkes and Docherty 1999; Helgason, Stölten & Engstrand 2003; Gordeeva and Scobbie 2007; Watson 2007; Clayton 2017), he reports more frequent productions and longer duration of pre-aspirated tokens among female speakers (Morris 2010: 7). This, he suggests, may be either due to physiological reasons (for more details see e.g. Helgason 2002: 229–231) or because his female participants were more regular users of Welsh (Morris 2010: 16). More recently, both Iosad (Forthcoming) and Spooner (2016) have investigated the devoicing of plosives following stressed vowels in south-eastern dialects of Welsh. In both these studies, pre-aspiration was noted as being a feature of fortis plosives.

Pre-aspiration has also been found in an in-depth study of Welsh English in Aberystwyth (Mid Wales). Hejná (2015) reported that pre-aspiration was on the whole a variable feature subject to both linguistic and extra-linguistic influences. Firstly, pre-aspiration was found to be near categorical only in the youngest speakers (98% in the data elicited from the youngest speaker, who was 24 years old in 2013). Secondly, although differences were found in the frequency of pre-aspiration between older female and male speakers, no such differences were found in the speech of the younger speakers (Hejná 2015: 261). These findings are highly relevant for a study of Welsh pre-aspiration in light of the fact that all of the participants were Welsh–English bilinguals and that linguistic features from Welsh are known to influence Welsh English either as substrate features or due to transfer in the speech of bilingual speakers (Mees & Collins 1999: 185–186; Thomas 1994, 1997; Paulasto 2006; Wells 1982: 351). A number of linguistic factors have also been reported by Hejná (2015): pre-aspiration and breathiness occurrence and duration are sensitive to the phonological length of the vowel, its phonological height, the place of articulation of the post-tonic obstruent, the manner of articulation of the post-tonic obstruent, and also stress.

1.4 Research questions

In order to provide a detailed account of pre-aspiration in Bethesda Welsh, we address the following research questions:

RQ1: Is pre-aspiration before fortis plosives a feature of the Welsh spoken in Bethesda, and can it also occur in the lenis plosives?

RQ2: To what extent do linguistic factors influence the presence of pre-aspiration, its duration and its noisiness, and do they do so in the ways reported for pre-aspiration in other languages?

RQ3: To what extent do extra-linguistic factors (speaker sex and home language) influence the presence of pre-aspiration, its duration and its noisiness, and how does this correspond to previous studies of sociophonetic variation in Welsh?

2 Community and Methodology

2.1 Community

The data were collected as part of a wider study which compared phonetic and phonological variation in north-west and north-east Wales (Morris 2013). The aim of that study was to examine the influence of linguistic factors as well as speaker sex, home language, and speech context (wordlist versus sociolinguistic interview) on the production of /r/ and /l/ in both the English and Welsh of bilingual speakers. It also compared differences in speech production between an area where Welsh is spoken by the majority of the local population (in the north west) and an area where English is the dominant community language (in the north east).

The decision to analyse a subset of these data was taken in order to reduce the extra-linguistic variables (namely speech context and area) in the analysis (see 2.3 below). Our aim in this study is to focus on the descriptive analysis of a linguistic feature of which there is very little previous description in an area which has not been considered in previous acoustic work. Indeed, there are very few acoustic descriptions of northern varieties of Welsh more generally (see Ball 1981 for a review of earlier work; Morris 2013; Cooper 2015). In order to achieve this aim, we do not address areal, contextual, or cross-linguistic variation so that (1) comparisons can be made with similar studies of pre-aspiration across languages and (2) further community-specific studies can build on this work in order to ascertain the extent to which pre-aspiration is a linguistic feature between different varieties of Welsh.

The subset analysed in the current study come from the Welsh wordlist data elicited from 16 Welsh–English bilinguals from the town of Bethesda, north-west Wales. The town is located in Gwynedd, the county with the highest proportion of Welsh speakers, where 77.5% (n = 3501) of the population of the town speak the language (Welsh Language Commissioner 2016). In light of our aims to control for possible areal variation and concentrate on one community (see above), Bethesda was chosen as the site of data collection (a local high school) did not have a large catchment area in comparison to other towns included in the corpus. Consequently, the speakers included in the dataset all lived in Bethesda or one of the surrounding villages.

2.2 Speakers

Speakers were aged between 16 and 18 years old at the time of collection. As stated above, speakers were recruited from a high school which serves the local area and all speakers had been born in Bethesda and the surrounding area or had moved there before starting formal education. At the high school, both pupils from Welsh-speaking and non-Welsh-speaking homes are taught together primarily in Welsh. The majority of pupils at the school come from Welsh-speaking homes. The sample of 16 speakers was stratified equally according to sex and home language (Welsh or English). This yielded four speakers per cell.

2.3 The wordlist

The decision to include wordlist data only has consequences given the emphasis in variationist sociolinguistics on vernacular speech. However, the comparison of sociolinguistic interview and wordlist data as a proxy for level of formality is in itself problematic (e.g. Coupland 1980). As stated in Section 2.1, the current study includes only wordlist data in order to minimise any potential effects of contextual variation while examining the influence of linguistic factors and the extra-linguistic factors of sex and home language.

The wordlist contained a total of 200 words which were presented on a laminated sheet of A4 paper without a carrier phrase. Of these 200 words, 55 unique words which contained a plosive were presented. We extracted and analysed each token which contained a plosive, including some instances where either intentionally (n = 32) or unintentionally (n = 2) participants repeated the same word twice, (n = 34). This yielded 914 tokens. A total of 39 tokens were excluded as we were unable to make reliable acoustic measurements. This meant that a total of 875 tokens were included in the initial analysis (see section 3.1).

Table 1 shows the number of tokens and the number of unique words (in brackets) for each word position in the fortis and lenis series. Tables A1 and A2 in the Appendix show the full list of words included in the current study.

Table 1: Number of tokens (and unique words) in each word position in the fortis and lenis series.

Consonant Series	**Word-medial**	**Word-final**	***Total***
Fortis	405 (24)	242 (15)	*647 (39)*
Lenis	188 (12)	40 (4)	*228 (16)*
Total	*593 (36)*	*282 (21)*	*875 (55)*

The word-medial tokens were all disyllabic and the word-final tokens were all monosyllabic. With some exceptions, stress falls on the penultimate syllable in Welsh (Hannahs 2013: 41) and therefore all of the plosives analysed for the current study occurred in stressed syllables.

Although all lenis and fortis plosives are permissible word-medially and word-finally (Awbery 1984: 81), the wordlist did not contain equal numbers of tokens containing /p/, /t/, /k/, /b/, /d/, /g/ in these word positions. In order to improve the statistical modelling, we grouped the plosives according to bilabial (*n* = 201), coronal (*n* = 418) and velar places of articulation (*n* = 256), whilst also keeping the variable of consonant series (fortis and lenis).

All of the tokens in the dataset contain monophthongs. Following Mayr & Davies (2011), the monophthong inventory for northern Welsh can be broadly described in terms of the short vowels /ɪ, ɨ, ʊ, ɛ, ə, ɔ, a/ and the long vowels /iː, ɨː, uː, eː, oː, aː/. Note that the unpaired /ə/ vowel can appear in stressed syllables but does not occur in monosyllabic content words (Hannahs 2013: 26). In monosyllabic (stressed) words, vowels are phonologically short preceding fortis plosives and long preceding lenis plosives (Awbery 1984: 66). In north-west Wales, stressed penultimate syllables (such as those found in the wordlist) are always short. This differs from southern dialects (where short vowels are found before fortis plosives and long vowels are found before lenis plosives) and central and north-eastern areas where variation between long and short vowels is thought to occur (see Awbery 1984 for an overview). The wordlist did not contain an equal number of tokens for each vowel category. For instance, some vowel categories appeared in more words than others, there are no examples of plosives following /ə/ in the dataset, and only monosyllabic tokens containing lenis plosives were preceded by long vowels.

In order to improve the robustness of the statistical modelling and in light of the results showing correlations between vowel height and backness with pre-aspiration (see Section 2.6), we grouped some vowels together for the analysis. The vowel categories included in the analysis were high front (/ɪ/ /ɨ/, /iː/), mid front (/ɛ/), low front (/a/), high back (/uː/, /ʊ/), and mid back (/ɔ/, /oː/). Vowel duration was analysed as a separate factor in the analyses (see Section 2.6). The total number of tokens in each vowel category are shown in Table 3 in Section 3.2.

2.4 Data collection

Data were collected in a quiet room on school premises. Participants were asked to read each word clearly at a natural speed. Participants were recorded reading the wordlist using a ZoomH2 recorder attached to an omnidirectional Audio-Technica Lavalier (lapel) microphone with 50 – 18,000 Hz frequency response and -54 dB sensitivity. All sessions were recorded in WAV format with a sampling frequency of 44.1 KHz and 16-bit quantisation. The recording of the wordlist took place following a sociolinguistic interview in Welsh (see 2.1 above) and lasted on average seven minutes. Data were transferred to a laptop computer and transcribed using ELAN (Sloetjes and Wittenburg 2008).

2.5 Acoustic analysis

We report our results separately for the two components traditionally assumed to form pre-aspiration in a broader sense: voiced breathiness and voiceless pre-aspiration (cf. Ní Chasaide 1985; Nance & Stuart-Smith 2013; Hejná 2015, 2016). Voiceless pre-aspiration and breathiness may show differences in phonetic and phonological conditioning, especially as breathiness (see Section 1.1).

This paper therefore focuses on three aspects of breathiness and pre-aspiration: frequency of occurrence, duration, and noisiness. Acoustic analyses were conducted in Praat (Boersma & Weenink 2015).

We decided to include lenis tokens in our analysis due to previous claims that voicing is a variable feature of lenis plosives in Welsh (see Section 1.2). In order to examine pre-aspiration before lenis plosives, we coded each token for voicing by visually inspecting the spectrogram and took voicing duration measurements (not reported here). The results of this analysis are included in Section 3.1.

2.5.1 Identifying breathiness and pre-aspiration

Each token was segmented and labelled as shown in Figures 1 and 2. The start of the preceding vowel was labelled at the onset of modal voicing. Breathiness (if applicable) was identified by a change to a quasi-sinusoidal waveform following the offset of modal voicing for the vowel and the onset of glottal friction as the glottis abducts (Nance & Stuart-Smith 2013: 130). The onset of pre-aspiration (if applicable) was marked at the point where voicing ceased, leaving voiceless glottal friction in the higher frequency portion of the spectrogram. The end of pre-aspiration was labelled at the end of glottal friction for the closure of the plosive. The end of the closure was labelled at the point of the plosive release.

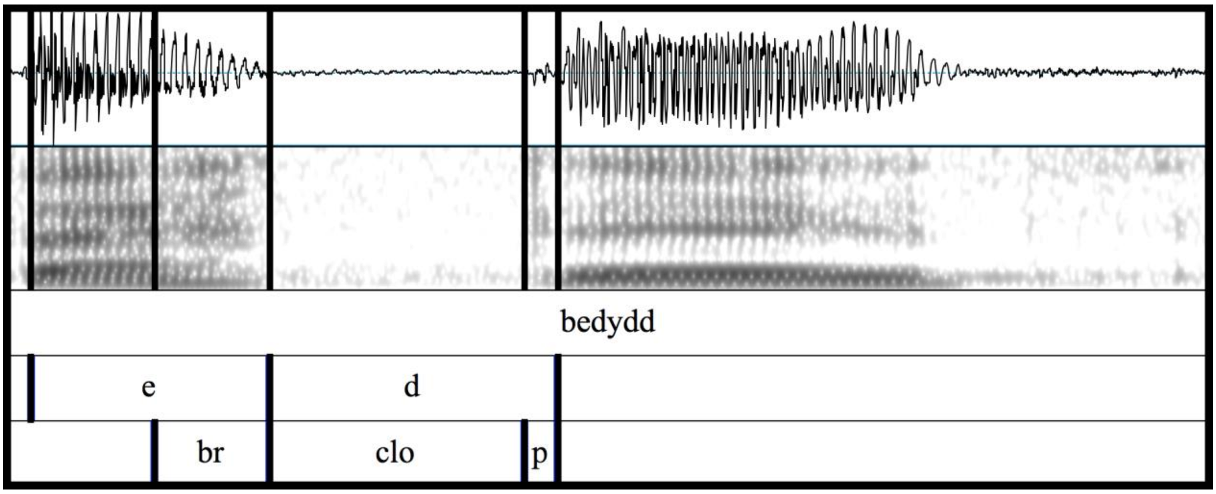

Figure 1: Example of the identification and segmentation of a token in word-medial position showing breathiness; br = breathiness; clo = closure; p = post-aspiration.

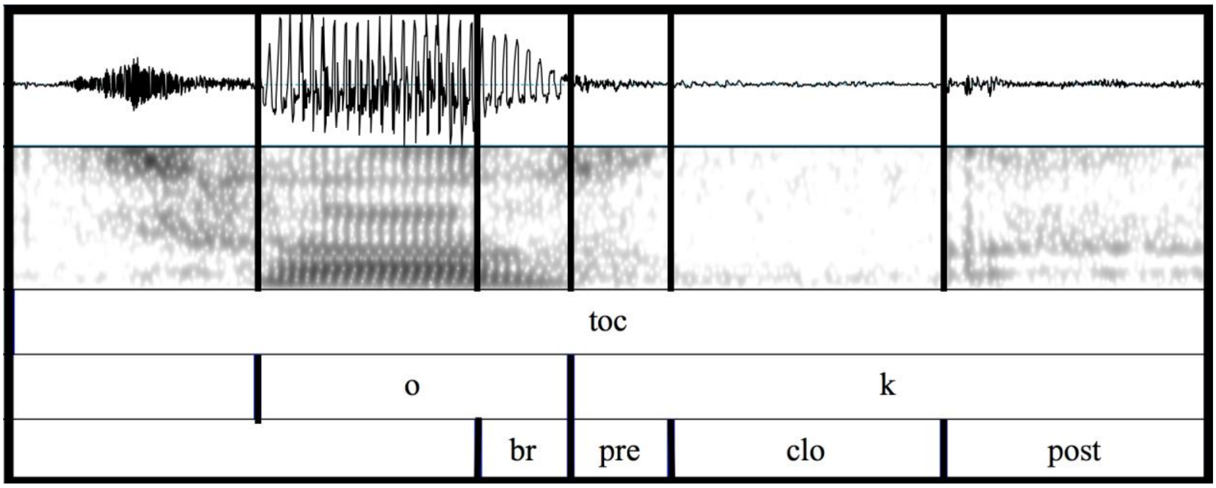

Figure 2: Example of the identification and segmentation of a token in word-final position showing breathiness and pre-aspiration; br = breathiness; pre = pre-aspiration; clo = closure; post = post-aspiration.

2.5.2 Quantifying breathiness and pre-aspiration

The analyses of duration are based on non-normalised durational data (ms). We originally included normalised data separately to the non-normalised data in the statistical modelling. This normalisation was carried out by converting the duration of pre-aspiration into the percentage of the overall duration of the word in question. The normalised data proved problematic, however, in comparing differences between pre-aspiration in word-medial and word-final contexts. Having compared the speech rate of participants impressionistically, we concluded that including non-normalised data in the final analysis would lead to more accurate results (see also Nance & Stuart-Smith 2013: 134). Durational measurements for both the preceding vowel and closure prior to the plosive release were extracted and included in the statistical modelling (see Section 2.6).

A relatively new measure, band-pass filtered zero crossing rate (BP ZCR), has been used in recent studies in order to measure the noisiness of breathiness and pre-aspiration (Gordeeva & Scobbie 2010; Nance & Stuart-Smith 2013). BP ZCR calculates how many times per second the sound wave in a band-pass filtered signal crosses a certain point (zero point) in both periodic and aperiodic periods of aspiration (Nance & Stuart-Smith 2013: 147). The measure gives the mean zero-crossings per second: the noisier a period of breathiness or pre-aspiration is, the more frequently the wave crosses zero within a time interval. This mid-frequency noise is an important cue for the perception of breathiness and pre-aspiration before voiceless plosives (Nance & Stuart-Smith 2013: 130).The durational and BP ZCR measures were extracted with scripts designed by Shigeto Kawahara (modified by James Kirby and Wendell Kimper) and Olga Gordeeva, after the manual segmentation of the material.

2.6 Statistical Analysis

For the statistical analyses of the presence of breathiness and pre-aspiration (a categorical dependent variable), mixed effects logistic regression analyses were conducted. For the analyses of duration and BP ZCR (continuous variables), we used mixed effects linear regression analyses. All statistical modelling was carried out using the lme4 (Bates, Maechler, Bolker & Walker 2016), lmerTest (Kuznetsova 2015), and Effects (Fox et al. 2016) packages for R (R Development Core Team 2015) in RStudio (2015). Mixed effects modelling was chosen primarily as models can (1) deal with both binary and continuous data, (2) consider random effects such as speaker or item, and (3) are relatively robust when dealing with unbalanced datasets (see, for example, Baayen 2008).

Table 2 shows the predictor variables included in each analysis. Italicised factor levels indicate the level taken to be the baseline. A step-up approach to the statistical modelling was taken whereby variables were added to the model and discarded if they did not significantly improve the model fit. Similarly, interactions were added to the model and removed if they did not result in an improved model. The models were compared using log-likelihood tests and deemed to be improved when there was a significant difference between the models and a reduction in the model deviance. In each model, speaker and lexeme were treated as random intercepts. We also attempted to add word, closure duration and vowel duration as random slopes in the analyses but these models failed to converge. The final (and best-fitting) models, along with the factors included in those models, are reported in Section 3 and can be found in Tables A3 – A8 in the Appendix.

Table 2: Independent variables considered in the statistical modelling.

Independent Variable	**Levels**	**Rationale**
Closure duration	Continuous variable	Longer closure duration has been found to be correlated with the shorter durations of pre-aspiration (e.g. Stevens & Reubold 2014: 455). This was not found in a recent study of Welsh English (Hejná 2015: 192).
Consonant Series	*Fortis* Lenis	Pre-aspiration in the fortis series may be an important correlate of the fortis-lenis contrast (e.g. Ní Chasaide & Ó Dochartaigh 1984: 151-2).
Speaker's Home Language	*English* Welsh	Previous studies of long-term societal bilingualism across languages have shown age of acquisition can influence speech production (e.g. Guion 2003; Simonet 2010). In Welsh-English bilingual speech, comparisons of phonetic variation in the production of monophthongs (Mayr et al. 2015) and /l/ (Morris 2017) have shown no such differences.
Place of Articulation	*Bilabial* Coronal Velar	Pre-aspiration is said to be most frequent and longest when preceding velar plosives and the least frequent and shortest when preceding bilabial plosives (Hejná 2015: 102).
Speaker Sex	*Female* Male	Female speakers generally produce more frequent and longer instances of pre-aspiration (e.g. Foulkes & Docherty 1999: 66; Gordeeva & Scobbie 2007: 17; Helgason, Stölten & Engstrand 2003).
Vowel Category	*Low front* Mid front High front Mid back	Studies have found that breathiness and pre-aspiration are more frequent and longer following lower vowels (Stevens & Hajek 2004b: 59; McRobbie-Utasi 1991).
Vowel Duration	Continuous variable	Longer durations of pre-aspiration have been attested following longer vowels (Clayton 2010), although the opposite holds for phonological vowel length.
Word Position	*Word-medial* Word-final	Hejná (2015: 115) found that both breathiness and pre-aspiration were longer word-finally in the English of Aberystwyth (mid Wales).

3 Results

In the following section, we address the extent to which voicing was found between the fortis and lenis plosives in both word-medial and word-final positions. We then examine the frequency of occurrence, duration, and BP ZCR measurements of breathiness and voiceless pre-aspiration.

3.1 Presence of voicing in the dataset

We firstly examined the presence of partial or full voicing in the production of the plosive (see Section 1.2). Of the 875 tokens, partial voicing of the plosive was found in 21.1% (*n* = 48/228) of the lenis tokens and in 0.15% of fortis tokens (*n* = 1/647). There were no instances of fully voiced plosives. Unsurprisingly, there were no instances of either breathiness or voiceless pre-aspiration in tokens where voicing was found in the consonant. Voiced examples of lenis tokens were found in the speech of all speakers with the exception of one. Figure 3 shows the percentage of voiced and unvoiced tokens in the fortis and lenis series for individual speakers.

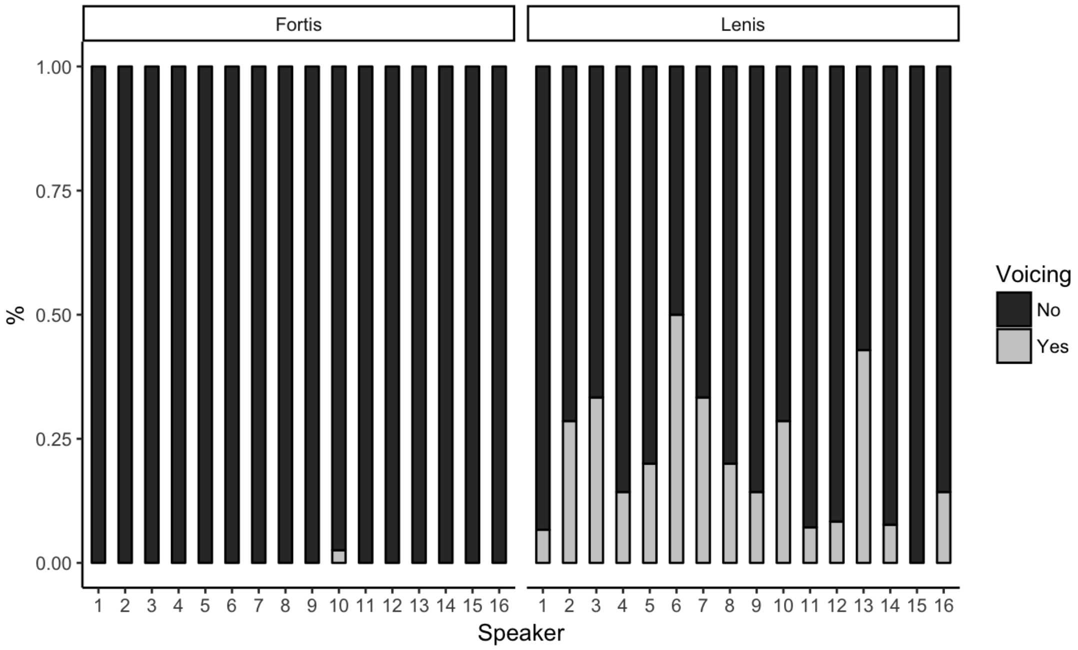

Figure 3: Percentage of voiced and voiceless tokens in fortis and lenis contexts by individual speaker (*n* = 875).

We also examined whether the presence of voicing is affected by the place of articulation of the lenis consonant: bilabial, coronal, and velar. The results of a Pearson's Chi-square test suggest that there is no significant difference between the presence of voicing in bilabial and coronal plosives ($\chi^2(1, N = 197) = 0.01$, $p = 0.93$). The differences between bilabial and velar plosives ($\chi^2(1, N = 88) = 1.09$, $p = 0.30$) and coronal and velar plosives

($\chi^2(1, N = 175) = 1.03$, $p = 0.31$) were also found not to be significant (Yates' correction for continuity was used in the latter two tests as there were four instances of voiced velar lenis plosives).

We did not include voiced tokens in the following analyses, which are based on tokens which contained an unvoiced plosive ($n = 826$). The reasons for this were that voiced tokens were relatively infrequent, appeared in most speakers' data, and were not sensitive to place of articulation effects regarding the frequency of application. Voiceless lenis tokens ($n = 180$) appeared in both word-medial contexts between vowels and word-finally.

In order to assess whether there were differences in length between the fortis and lenis series, particularly in word-medial position where fortis plosives are expected to be geminated (see Section 1.2), we conducted independent samples t-tests on the voiceless plosives in word-medial contexts and word-final contexts. In word-medial contexts, there was no significant difference between the duration of fortis ($M = 409.29$ms, $SD = 86.70$) and lenis plosives ($M = 414.03$ms, $SD = 71.7$, $p = 0.57$). In word-final contexts, there was also no significant difference between the duration of fortis ($M = 248.28$ms, $SD = 69.68$) and lenis plosives ($M = 248.48$ms, $SD = 61.17$, $p = 0.91$). This suggests that any differences found between fortis and lenis plosives in word-medial and word-final contexts will not be due to durational differences between the two series[1].

3.2 Frequency of breathiness and pre-aspiration

In the subset of tokens which did not contain a phonetically voiced plosive ($n = 826$), a period of breathiness was found in 74.3% ($n = 614/826$) of tokens. Voiceless pre-aspiration was found in 72.2% (n = 596/826) of tokens. Of the total number of unvoiced tokens, 66.2% ($n = 547/826$) contained both breathiness and pre-aspiration.

In the same subset of tokens which did not contain a voiced plosive, a period of breathiness was found in 79.6% ($n = 514/646$) of the fortis tokens compared to 55.6% ($n = 100/180$) of lenis tokens. Voiceless pre-aspiration was found in 79.7% ($n = 515/646$) of the fortis tokens compared to 44.5% ($n = 81/180$) of the lenis tokens. The data from individual speakers show that voiceless pre-aspiration is a variable feature preceding voiceless fortis plosives for all speakers, with the exception of Speaker 8, who pre-aspirated all fortis tokens ($n = 40$). Pre-aspiration preceding lenis tokens was found in the speech of all participants with the exception of Speakers 6 and 7. Figure 4 shows the percentage of pre-aspirated tokens in voiceless fortis and lenis contexts for individual speakers.

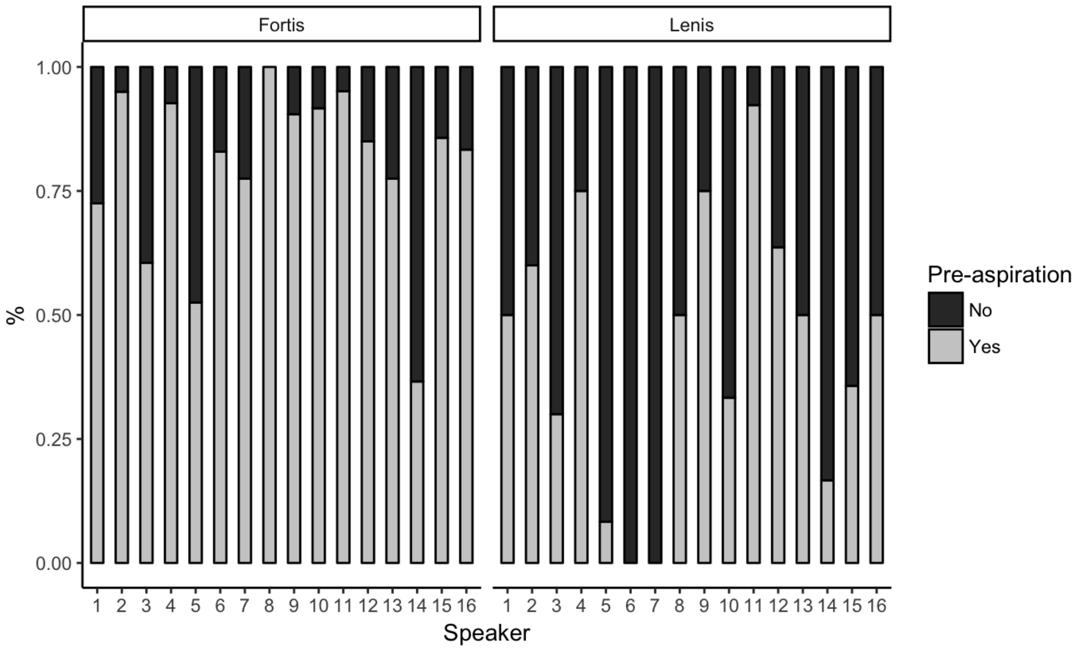

Figure 4: Percentage of pre-aspirated and non-pre-aspirated tokens in voiceless fortis and lenis contexts by individual speaker (*n* = 826).

The effects of linguistic and extra-linguistic factors on the presence of (1) breathiness and (2) voiceless pre-aspiration were examined using mixed effects logistic regression modelling (see Section 2.6). Separate models were run with either breathiness or pre-aspiration as response variables and, in each model, both speaker and lexeme were coded as random factors. The response variables tested were speaker sex, home language, preceding vowel, place of articulation, consonant series, word position, vowel duration (ms) and closure duration (ms). The best-fitting models for both breathiness and voiceless pre-aspiration contained preceding vowel, place of articulation, consonant series, and speaker sex as predictor variables. An interaction between place of articulation and consonant series was retained in the final models. These factors were not found to be significant but significantly improved the model fit for both breathiness (χ^2 (5) = 33.75, $p < 0.01$) and pre-aspiration (χ^2 (5) = 59.02, $p < 0.01$) when compared using log-likelihood tests (see Section 2.6). As there was no breathiness or pre-aspiration in the phonetically voiced tokens, and due to the fact that voiced tokens appeared with relatively low frequency in the speech of most participants, they were not included in the final model. The total number of tokens included in the statistical modelling was 826.

The full results of the statistical modelling can be seen in Tables A3 and A4 in the Appendix. Consonant series was not found to be a significant predictor on the presence of breathiness at an alpha level of 0.05 (β = -1.51, z = -1.87, p = 0.06). The presence of voiceless pre-aspiration, however, was less likely before voiceless lenis plosives when compared to fortis plosives (β = -2.37, z = -3.17, p < 0.01).

In addition to consonant series, place of articulation was found to be a significant predictor. Breathiness was more likely to be present before both coronal (β = 1.87, z = 3.35, p < 0.01) and velar plosives (β = 1.87, z = 3.26, p < 0.01) than bilabial plosives. Of the voiceless tokens containing a velar plosive, 83.3% (n = 210/252) contained breathiness compared to 77.2% (n = 298/386) of tokens containing a coronal plosive and 56.4% (n = 106/188) of tokens containing a bilabial plosive. Pre-aspiration was also more likely to be present preceding the coronal (β = 1.54, z = 3.21, p < 0.01) and velar (β = 2.82, z = 5.2, p < 0.01) plosives than when preceding bilabial plosives.

Of the tokens containing a velar plosive, 89.3% (n = 225/252) contained pre-aspiration compared to 71.5% (n = 276/386) of tokens containing a coronal plosive and 50.5% (n = 95/188) of tokens containing a bilabial plosive. Figures 5 and 6 show the percentage of tokens found to contain breathiness and pre-aspiration respectively grouped by both consonant series and place of articulation for comparison.

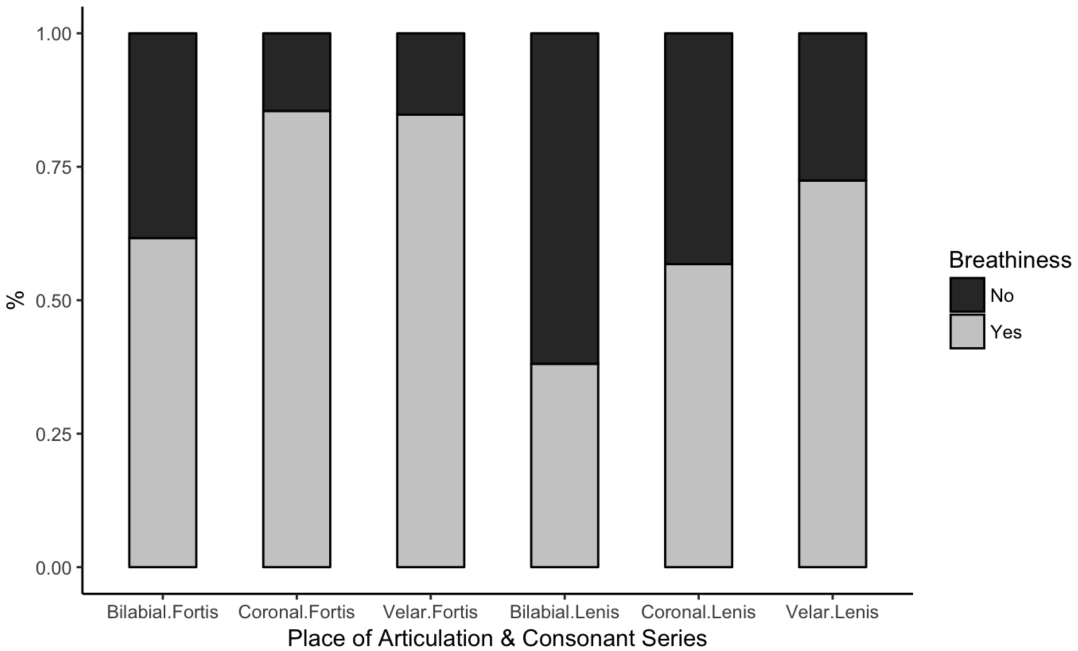

Figure 5: Percentage of breathy and non-breathy voiceless tokens by place of articulation and consonant series (n = 826).

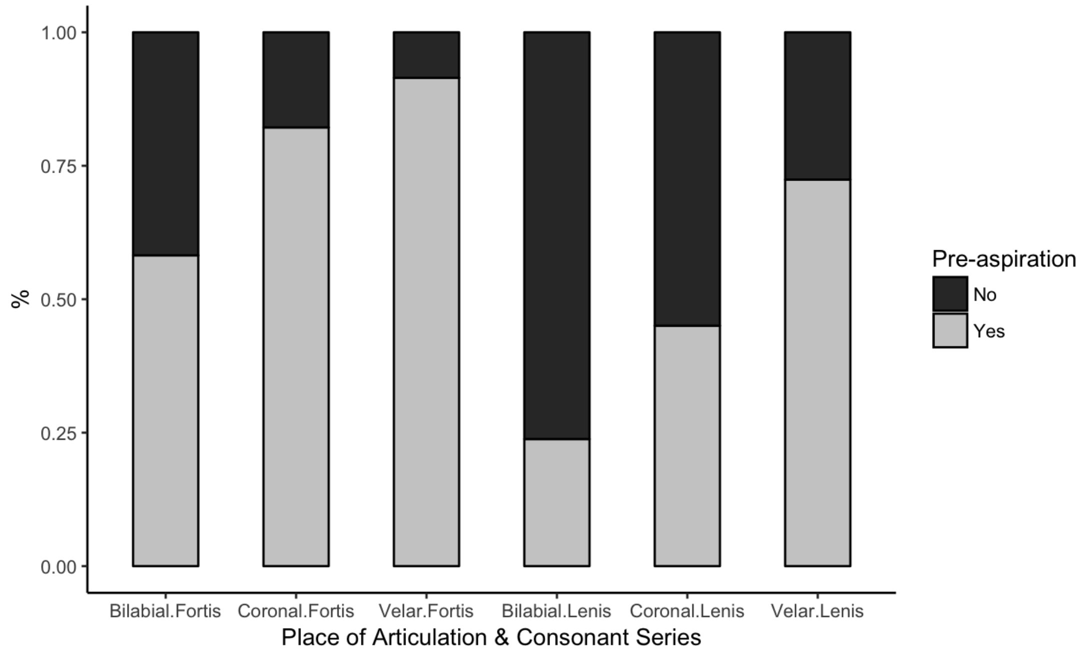

Figure 6: Percentage of pre-aspirated and non-pre-aspirated voiceless tokens by place of articulation and consonant series (*n* = 826).

The preceding vowel was found to be a significant predictor of both breathiness and pre-aspiration in the models. When compared to low front vowels as the baseline, breathiness was found to be less likely to occur following high back vowels (β = -1.34, z = -2.37, p = 0.02). Compared to the same baseline vowel category, pre-aspiration was less likely to occur following high front vowels (β = -1.25, z = -2.66, p < 0.01), mid back vowels (β = -1.13, z = -2.23, p = 0.03), and high back vowels (β = -1.84, z = -3.65, p < 0.01). Table 3 shows the percentage of breathy and pre-aspirated tokens following each vowel. The fact that breathiness following high front vowels was not found to be significantly less likely than when following low front vowels is interesting given the descriptive data in Table 3. This may be due to the relatively low number of tokens in each vowel category and should be treated with caution.

Table 3: Percentage and number of breathy and pre-aspirated voiceless tokens by vowel category (*n* = 826).

Vowel	**Breathiness**		**Pre-aspiration**	
	%	*n* (total)	%	*n* (total)
Low front	82.9	242 (292)	82.5	241 (292)
Mid front	77.0	107 (139)	92.8	109 (139)
Mid back	72.3	94 (130)	67.7	88 (130)
High back	65.4	72 (110)	60.0	66 (110)
High front	63.9	99 (155)	59.4	92 (155)

The results of the statistical modelling also suggest that male speakers are less likely to produce both breathiness (β = -1.42, z = -2.33, p = 0.02) and pre-aspiration (β = 1.78, z = -

3.14, $p < 0.01$). Of the voiceless tokens produced by male speakers, 64.7% (n = 269/416) contained breathiness and 61.8% (n = 257/416) contained voiceless pre-aspiration. Of the tokens produced by female speakers, 84.1% (n = 345/410) contained breathiness and 82.7% (n = 339/410) contained voiceless pre-aspiration.

3.3 Duration

In order to examine the duration in the dataset, we conducted mixed effects linear analyses on both the non-normalised and normalised measurements with (1) duration of breathiness, (2) duration of voiceless pre-aspiration, and (3) duration of both breathiness and voiceless pre-aspiration as separate response variables. Again, we excluded the phonetically voiced tokens (n=49/875) from the analysis as there were no instances of either breathiness or voiceless pre-aspiration. The total number of tokens included in the statistical modelling was 826. As the factors influencing duration were different for breathiness compared to those influencing voiceless pre-aspiration, we report the results separately here. As in the models for frequency of occurrence, both speaker and lexeme were added to the models as random intercepts. The predictor variables tested were speaker sex, home language, preceding vowel, place of articulation, consonant series, word position, vowel duration (ms) and closure duration (ms). Interactions between place of articulation, consonant series, closure duration, and preceding vowel were also tested.

The full model for breathiness is shown in Table A5 in the Appendix. The best-fitting model for breathiness contained place of articulation, consonant series, word position and closure duration as predictor variables. The results of the model indicate that the duration of breathiness is significantly shorter preceding lenis plosives than when preceding fortis plosives ($\beta = -6.56$, $t = -2.12$, $p < 0.01$). The mean duration of breathiness was 21.09ms (SD = 14.49) preceding fortis plosives compared to 12.26ms (SD = 15.37) preceding lenis plosives.

Place of articulation is also a significant predictor with breathiness tending to be longer preceding coronals ($\beta = 5.74$, $t = 2.40$, $p = 0.02$). The mean duration of breathiness preceding coronals was 21.86ms (SD = 15.55). For velars, the mean duration of breathiness was 20.49ms (SD = 12.99). For bilabials, the mean duration of breathiness was 14.86ms (SD = 15.16). Figure 7 shows the duration of breathiness grouped by consonant series and the place of articulation of the preceding consonant.

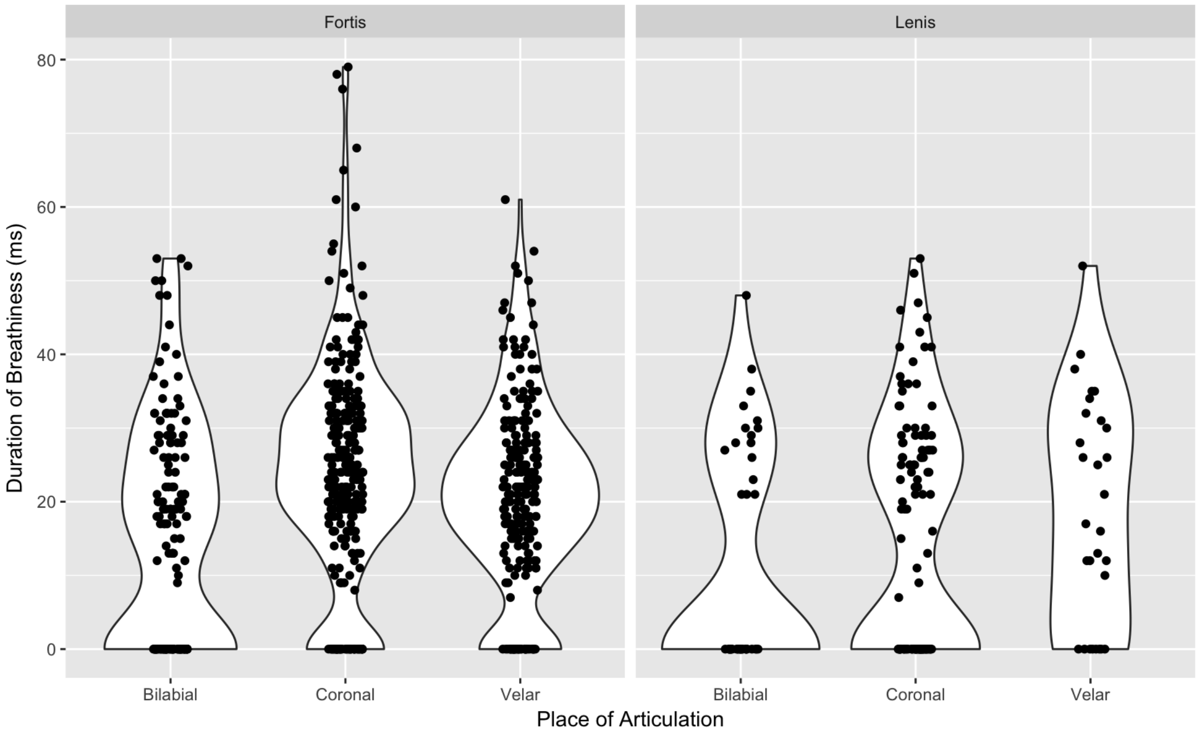

Figure 7: Duration of breathiness (ms) by place of articulation and consonant series (*n* = 826).

The best-fitting model for the duration of pre-aspiration (ms) contained speaker sex, home language, preceding vowel, place of articulation, and consonant series. The full model can be found in Table A6 in the Appendix. Interactions were removed during the modelling as they were not found to be significant and did not significantly improve the model.

The results indicate that the duration of pre-aspiration is significantly shorter preceding lenis tokens than preceding fortis tokens (β = -16.61, t = -2.56, p = 0.01). The mean duration for fortis tokens was 29.50ms (*SD* = 23.9) compared to 12.38ms (*SD* = 18.04) for the lenis tokens.

Pre-aspirated tokens in velar contexts were also more likely to be longer compared to bilabial contexts (β = 12.53, t = 2.84, $p < 0.01$), which differs from the results for breathiness (see above). The mean duration of pre-aspirated tokens in velar contexts was 34.64ms (*SD* = 22.42). The mean pre-aspiration for tokens preceding bilabial plosives was 18.19ms (*SD* = 24.27). The mean duration of pre-aspirated tokens in coronal contexts was 23.58ms (*SD* = 22.77), although this was not found to be significantly different for the duration of pre-aspiration preceding bilabials. Figure 8 illustrates the mean duration of pre-aspiration (ms) by consonant series and place of articulation.

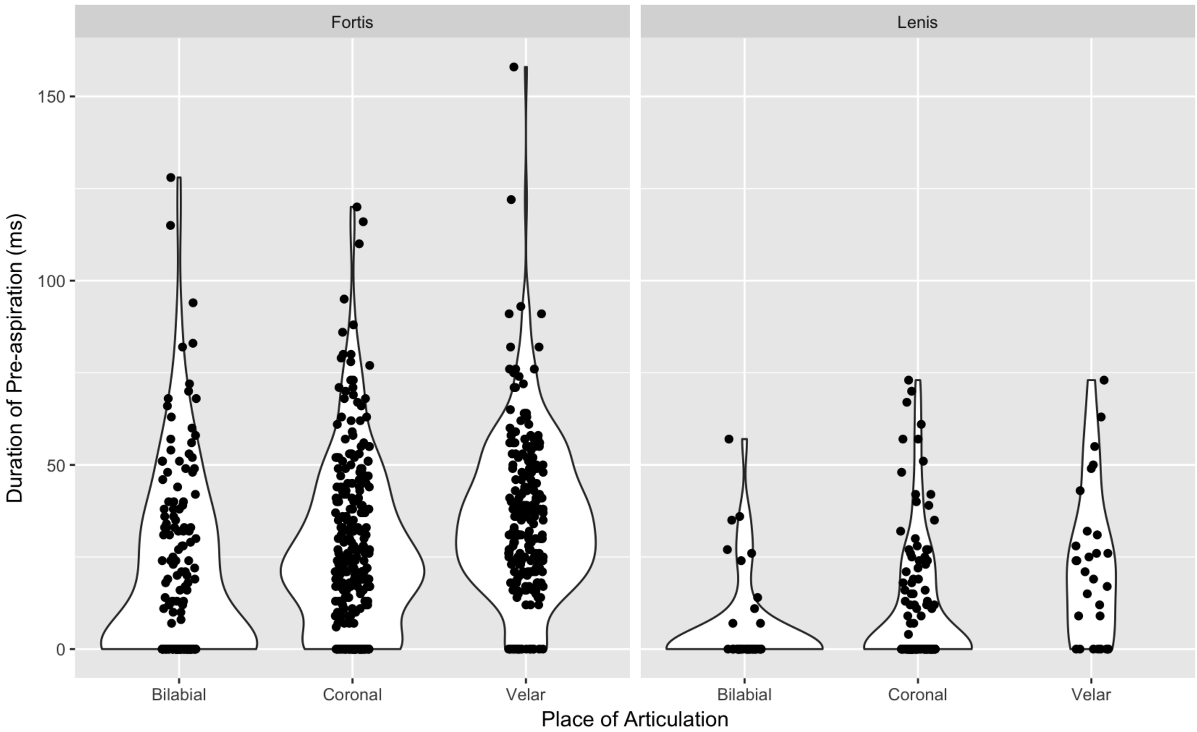

Figure 8: Duration of pre-aspiration (ms) by place of articulation and consonant series (n = 826).

Unlike in the results for the duration of breathiness, some vowel categories were found to be significant when compared to the baseline vowel of /a/. Specifically, voiceless pre-aspiration was more likely to be shorter following high front (β = -13.01, t = -3.19, $p < 0.01$) and high back vowels (β = -15.28, t = -3.48, $p < 0.01$) when compared to the baseline. The average duration of voiceless pre-aspiration is shown in Figure 9:

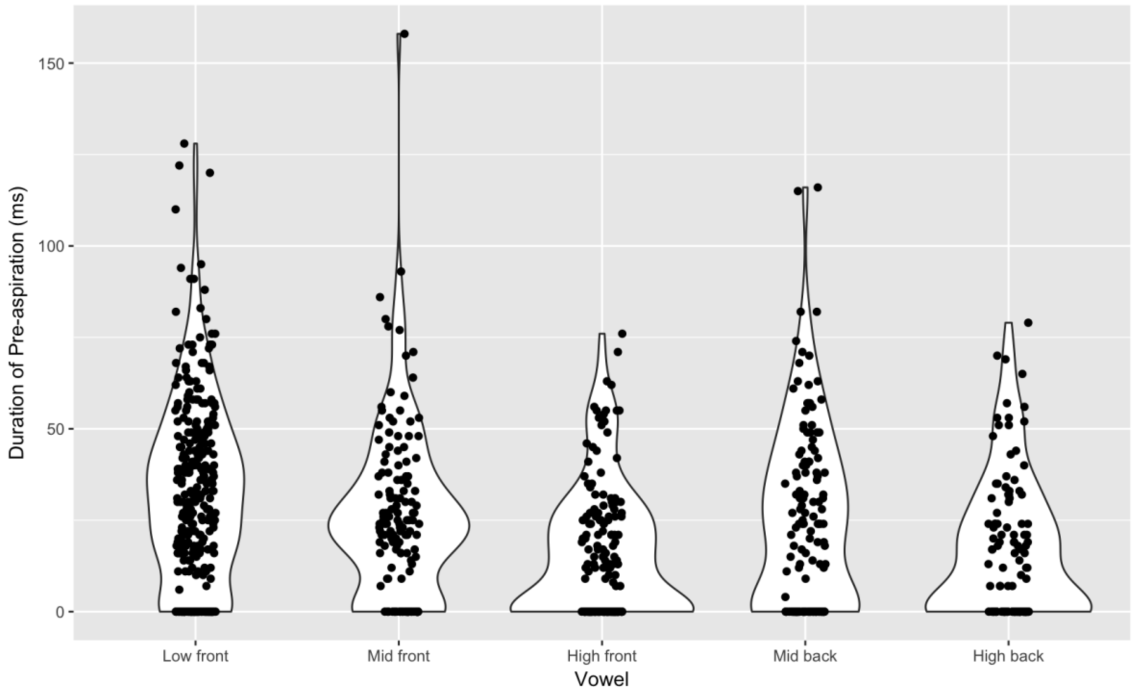

Figure 9: Duration of pre-aspiration (ms) by place of vowel category (n = 826).

The duration of pre-aspiration tended to be shorter in the data from male speakers than in the data from female speakers (β = -11.27, *t* = -2.81, *p* = 0.01). The mean duration of voiceless pre-aspiration for male speakers was 20.02ms (*SD* = 22.04). For female speakers, the mean duration of pre-aspiration was 31.52ms (*SD* = 24.2). Home language was also found to be significant in the statistical modelling, with those from English-speaking homes tending to produce shorter durations of pre-aspiration (β = -8.99, *t* = -2.24, *p* = 0.04). The mean duration of voiceless pre-aspiration among those from English-speaking homes was 21.31ms (*SD* = 19.45) compared to 30.03ms (*SD* = 26.78) among those from Welsh-speaking homes.

3.4 Band-pass filtered zero crossing rate (BP ZCR)

The BP ZCR measurements yield the mean zero-crossings per second split across five equal sections. A series of mixed effects linear analyses were conducted on the BP ZCR measurements. These included analyses of breathiness and pre-aspiration separately, as well as measurements taken from the offset of modal voicing to the plosive closure. In order to aid comparison with the frequency and durational data above, we report the analyses on breathiness and voiceless pre-aspiration separately. In line with the previous analyses, the results reported here concern only the phonetically voiceless tokens (*n* = 826). As we are not concerned with changes to zero-crossing rate over time, we base our analyses on the mean values.

As in the previous models, both speaker and lexeme were added to the models as random intercepts. The response variables tested were speaker sex, home language, preceding vowel, place of articulation, consonant series, word position, vowel duration (ms) and closure duration (ms). Interactions between place of articulation, consonant series, and preceding vowel were also tested. The best-fitting models for both breathiness and pre-aspiration included consonant series, place of articulation, vowel category, and speaker sex.

The models suggest that breathiness and pre-aspiration preceding lenis plosives are less noisy than when preceding fortis plosives (see Tables A7 and A8 for coefficients and full models). Both breathiness and pre-aspiration were found to be noisier when preceding coronal plosives and velar plosives when compared to bilabials. Figures 10 and 11 show the BP ZCR measurements for by consonant series and place of articulation of breathiness and pre-aspiration respectively.

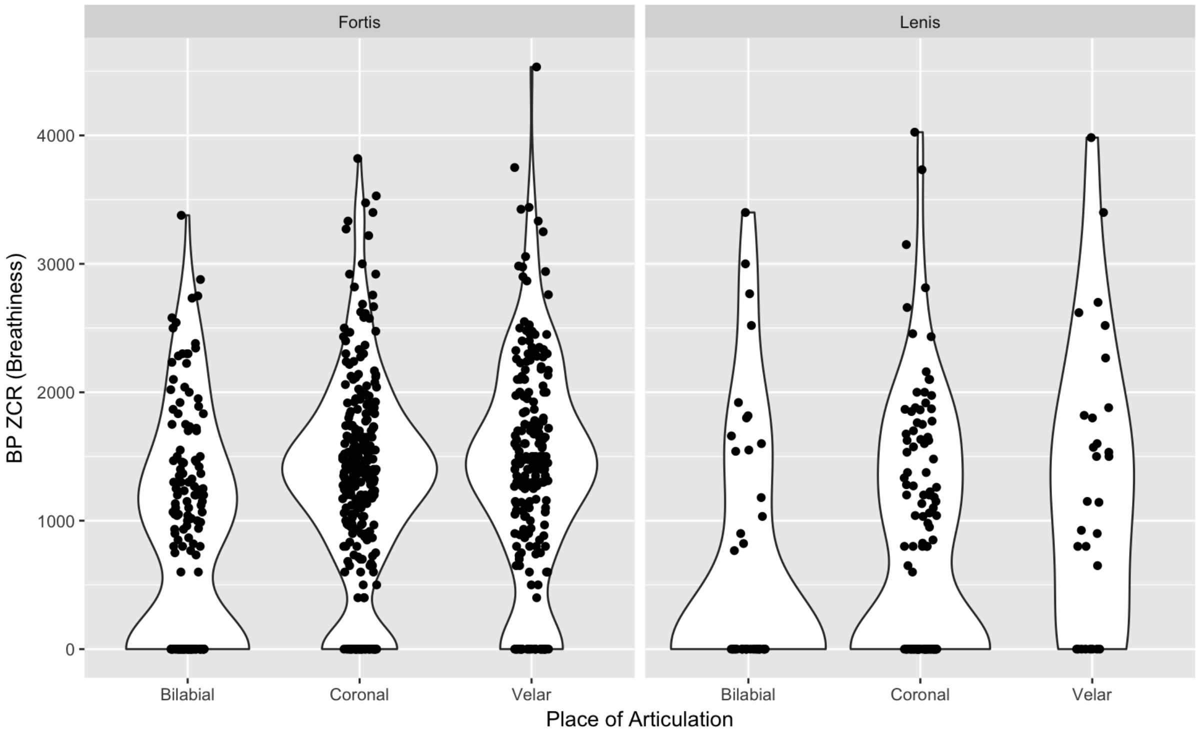

Figure 10: BP ZCR of breathiness by place of articulation and consonant series (*n* = 826).

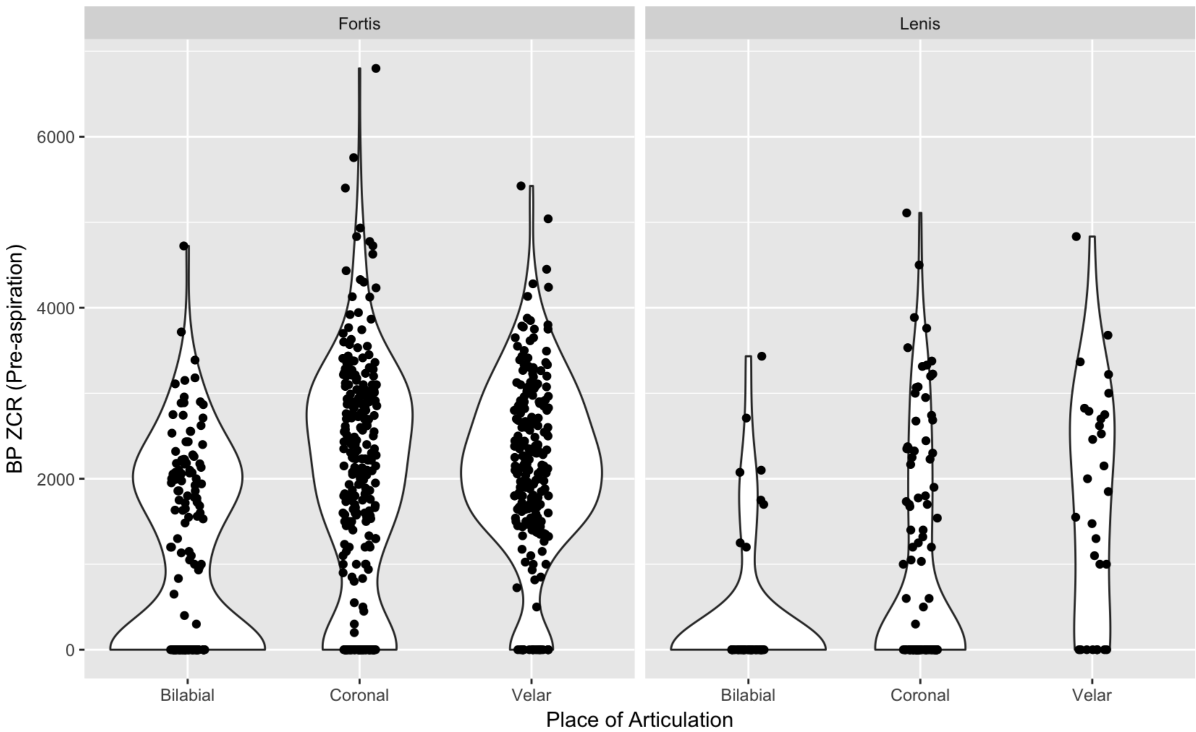

Figure 11: BP ZCR of pre-aspiration by place of articulation and consonant series (*n* = 826).

Vowel category was also found to be significant for the BP ZCR measurements for both breathiness and pre-aspiration. For breathiness, the mid front, high front, mid back and high back categories were all found to be less noisy when compared to the baseline category of low front. For pre-aspiration, the high front, mid back and high back vowels were found to be less noisy.

Finally, male speakers tended to produce less noisy instances of both breathiness (β = -538.90, t = -2.90, p = 0.01) and pre-aspiration (β = -760.96, t = -3.148, p < 0.01).

4 Discussion

The current study aimed to ascertain whether pre-aspiration is a feature of Bethesda Welsh. The results indicate that, for most speakers, it is a variable feature. This section further discusses the frequency of occurrence of pre-aspiration (RQ1) and both the linguistic (RQ2) and extra-linguistic factors (RQ3) which influence its occurrence, duration, and noisiness.

4.1 Frequency of occurrence of pre-aspiration

As stated above, pre-aspiration appears to be a variable feature and only one of the speakers produced all tokens as pre-aspirated in the fortis series (see Figure 5). Similarly, two speakers produced no pre-aspirated tokens in the lenis series. Pre-aspiration is not generally noted as a feature of the language in descriptions, yet we would argue that our results suggest that it is a feature which should be included in future studies of other dialects. This would allow areal comparisons to be made.

Pre-aspiration was also found in the voiceless lenis tokens though pre-aspirated tokens were significantly less likely in lenis contexts. Although comparisons between the fortis and lenis series are not wholly permissible due to the differing number of tokens, the fact that pre-aspiration was less frequent in the context of lenis tokens might suggest that it is a correlate of and possibly a cue to the fortis–lenis contrast in Welsh (as is the case for post-aspiration). Further speech perception tasks using controlled stimuli will shed light on this. However, we would argue that, as pre-aspiration is found in both contexts, it would be a more ambiguous correlate of the fortis–lenis contrast than post-aspiration, which is found in the fortis series only. Furthermore, the fact that pre-aspiration is a variable feature in voiceless fortis and lenis contexts diminishes its potential cue availability (in the sense of MacWhinney 2001, 2012).

The fact that pre-aspiration was found in lenis plosives is somewhat surprising in light of suggestions that the feature is a consequence of post-aspiration anticipation (e.g. Hejná 2015: 91, 265), and is therefore rare among lenis plosives (which typically lack post-aspiration). However, San Martín Itunyoso Trique (DiCanio 2012: 252-4) and Scottish Gaelic (Nance & Stuart-Smith 2013: 137-9) are notable examples of other languages that have been reported to show pre-aspiration both in the fortis and the lenis series. As the lenis plosives in the Bethesda data are not post-aspirated (despite being phonetically voiceless), we suggest that anticipatory accounts cannot explain the presence of pre-aspiration in the lenis series.

One possible scenario is that we may be dealing with an analogical extension from the fortis context in Bethesda Welsh and other varieties where pre-aspiration occurs preceding lenis consonants. Although pre-aspiration functions as a correlate of the fortis-lenis contrast in many pre-aspirating languages and their varieties (see Hejná 2015: chapter 6; for an overview), the phenomenon is not generally thought to be a particularly salient cue to this contrast (see Hejná & Kimper, Submitted) in comparison to other cues. The potential lack of perceptual salience (or an auditory negative bias) may be why pre-aspiration may be showing in both fortis and lenis series in Bethesda Welsh and the other two relevant languages.

4.2 Linguistic constraints on pre-aspiration

Cross-linguistically, pre-aspiration is subject to a number of linguistic constraints. This study focused on the place of articulation of the plosive and vowel quality and in both fortis and lenis contexts.

The results for place of articulation agree with the generally reported effects across languages: the bilabial plosives (/p, b/) were associated with the least frequent application of breathiness and voiceless pre-aspiration in comparison to the other places of articulation (/t, d/; /k, g/). The bilabial plosives also tended to have a shorter duration of breathiness and pre-aspiration, which is in agreement with previous work on pre-aspiration (Ní Chasaide 1985; Stevens & Hajek 2004a; Stevens & Hajek 2004b; Helgason & Ringen 2008; Nance & Stuart-Smith 2013; Clayton 2017). One exception is the duration of breathiness, where no differences were found between the bilabial and velar plosives.

Both pre-aspiration and breathiness also tended to be less noisy in the context of bilabial plosives as opposed to coronals and velars. Interestingly, voiceless pre-aspiration frequency and duration increase from the most anterior to the most posterior place of articulation in both series (/p, b/ > /t, d/ > /k, g/), which is not the case for breathiness (/p, b/ < /t, d, k, g/. It is not clear what the reasons are for the general patterning of bilabials to exhibit the least frequent and the shortest breathiness and pre-aspiration, as well as the least noisy breathiness and pre-aspiration. We suggest that differences in pre-aspiration properties across the different places of articulation may be affected by the articulator mass in contact. The production of post-aspirated /t/ and /k/, especially if also affricated (which is a feature of Welsh and Welsh English), requires for the contact between the two articulators to be greater than in the case of a post-aspirated and unaffricated /p/ (see Kang & Kochetov 2010 for an

EPG study of Korean plosives and affricates). This would lead us to expect greater differences in pre-aspiration across the places of articulation within the fortis series than in the lenis series, the latter of which is not associated with post-aspiration and affrication. However, at present there are no existing instrumental studies of affrication in Welsh, and testing whether articular mass can explain the patterns found remains to be addressed by future studies.

Previous work has shown that, across languages, low front vowels tend to be associated with more frequent and longer tokens of pre-aspiration as well as breathiness (e.g. Ní Chasaide 1985; McRobbie-Utasi 1992; Helgason 2002; Stevens & Hajek 2004a: Gordeeva & Scobbie 2007). The results from this study are inconclusive with regards to vowel quality. Both breathy and pre-aspirated tokens were less frequent following high back vowels when compared to low front vowels (the baseline) but no other vowel categories were found to be significant. The same holds for the duration of pre-aspiration (though not breathiness, where vowel category was not found to be significant).

Both breathiness and pre-aspiration were also found to be the noisiest with low front vowels in comparison to mid back and high vowels (pre-aspiration and breathiness) and mid front vowels (breathiness only). This positive correlation mirrors previous findings for frequency and duration (see above) though it is somewhat surprising as we would expect noisier friction in the context of high vowels due to the narrower oral constriction (Ohala 2005: 420). It remains to be seen whether the interaction between the oral cavity and the larynx may shed light on why low vowels are associated with the most frequent, longest, and noisiest pre-aspiration in pre-aspirating languages. However, it should also be noted that the upper limit of the filtering in the BP ZCR measurements is at 5,500Hz, which means that friction at higher frequencies (more akin to pre-affrication rather than pre-aspiration) is not taken into consideration. To fully understand the developments of glottal and oral pre-aspiration and pre-affrication, further articulatory studies should examine different states of the larynx associated with the articulation of different vowels and acoustic studies should consider higher frequencies when applying the BP ZCR measure in order to examine these claims further.

4.3 Extra-linguistic constraints on pre-aspiration

Male speakers tended to produce less frequent and less noisy instances of both breathiness and pre-aspiration. Pre-aspiration also tended to be shorter when produced by male speakers although the same is not valid for breathiness. These patterns correspond to the general tendency for females to be more prone to pre-aspirate and to pre-aspirate with longer durations found in other studies (see e.g. Hejná 2015: chapter 7) and have often been attributed to vocal-tract differences between men and women (e.g. Helgason 2002). This explanation is far from straightforward, however, especially considering vocal tract measurements are not taken as part of such studies. In addition, some studies have found few differences between male and female speakers' productions of pre-aspiration (e.g. Stevens 2010). Hejná (2015: 247–248) found no differences between younger generations of male and female speakers in her study of Aberystwyth English, which suggests that physiological differences may not be as linked to biological sex as previously assumed or that physiological differences can be overridden in instances where social factors become more salient. It is therefore unclear whether Welsh pre-aspiration is subject purely to physiological sex differences or also social gender differences.

Recent work on phonological and phonetic variation in this area of Wales has found differences in speech production between male and female speakers for which a more socially-motivated explanation might be more suitable. In his study of (r) variation, Morris (2013) found that female speakers were more likely than male speakers to produce the standard variant of /r/ (the voiced alveolar trill [r] rather than the alveolar approximant [ɹ]) in Welsh speech. Morris (2017) found that female speakers tended to produce lighter realisations of /l/ in Welsh than male speakers (where /l/ is expected to be dark in all syllable positions) but also differentiated to a greater extent than male speakers between their realisations of /l/ in Welsh and English.

A similar claim can be made with regards to the finding that speakers from English-speaking homes were more likely to produce shorter breathiness and pre-aspiration than those from Welsh-speaking homes (frequency and noisiness are not affected by home language.). Although previous work across languages has shown clear age of acquisition effects on speech production among bilingual speakers (Piske et al. 2001), research on Welsh-English bilingual speech has found home language to be insignificant in the production of monophthongs (Mayr et al. 2015) and /l/ (Morris 2017) in both English and Welsh. However, significant differences between home-language groups have been found in the production of /r/ in both languages in Gwynedd, and have been attributed to the tendency for linguistic background to be an important marker of peer-group membership in this community (Morris 2013).

This suggests that a more socially-oriented explanation for the sex and home language variation found for pre-aspiration should be investigated in further work which considers both the perceptions of this variable among speakers and possible socio-indexical meanings on the one hand, and social structures in the local community on the other hand.

5 Conclusion

This paper investigated pre-aspiration in the Welsh of Bethesda (Gwynedd). Specifically, it sought to examine the frequency, duration, and noisiness of pre-aspiration in the context of voiceless plosives.

It was shown that pre-aspiration is a variable feature in both fortis and, to a lesser extent, voiceless lenis contexts. The finding that the feature is also present before lenis plosives and the lack of voicing in non-initial contexts in the lenis series has hitherto been unreported.

We found that the presence of pre-aspiration, its duration, and its noisiness was influenced by linguistic factors which have also been shown to be significant in previous studies across languages. Specifically, pre-aspiration was less frequent, shorter, and the least noisy preceding bilabial plosives than when preceding coronal and velar plosives. In addition, pre-aspiration tended to be more frequent, longer, and noisier following front low vowels.

The influence of a speaker’s home language on pre-aspiration was not found to be consistent. Those from English-speaking homes were more likely to produce shorter pre-aspirated tokens, but the frequency and the noisiness of the phenomenon were not affected. Speaker sex was, however, found to be a consistently significant predictor across the three measures of pre-aspiration in the study in both fortis and lenis contexts. Male speakers were less likely to produce pre-aspirated tokens, and the duration of their pre-aspiration was shorter and less noisy than tokens produced by female speakers. The correlation between speaker sex and the three measures of pre-aspiration does support the results of previous studies across languages to a certain extent, and may be due to physiological differences between men and women.

Further acoustic and articulatory work should examine this feature using a more experimental design which allows for more control over vocalic context and which includes more speakers. Ethnographically-informed linguistic studies and perception experiments will also shed light on the extent to which pre-aspiration carries socio-indexical meaning for Welsh speakers and whether this might explain the differences found between male and female speakers and those from Welsh-speaking and English-speaking families.

Acknowledgements

The authors would like to thank anonymous reviewers for their helpful comments. Any remaining errors are, of course, our responsibility.

Appendix

Table A1: Word-medial tokens

	Phoneme	**Token (Welsh)**	**English**	**Preceding vowel (based on auditory inspection and Mayr & Davies 2011)**
1.	/p/	capel	chapel	/a/
2.		cwpan	cup	/ʊ/
3.		lapio	to wrap	/a/
4.		popeth	everything	/ɔ/
5.		tipyn	a little (repeated twice)	/ɪ/
6.	/t/	atal	to stop	/a/
7.		atat ti	towards you	/a/
8.		Betws	Bed-house or house of prayer (heard frequently as part of place names in Wales)	/ɛ/
9.		eto	again	/ɛ/
10.		Guto	Guy (name)	/ɨ/
11.		llety	accommodation	/ɛ/
12.		patrwm	pattern	/a/
13.		petawn	if I were	/ɛ/
14.		potel	bottle	/ɔ/
15.		pwtyn	stub	/ʊ/
16.		tatws	potatoes	/a/
17.	/k/	cacen	cake (repeated twice)	/a/
18.		lwcus	lucky	/ʊ/
19.		paced	packet	/a/
20.		pecyn	pack	/ɛ/
21.		picio	to pick	/ɪ/
22.		roced	rocket	/ɔ/
23.		taclus	tidy	/a/
24.		tocyn	ticket	/ɔ/
25.	/b/	baban	baby	/a/
26.		pobl	people	/ɔ/
27.		ruban	ribbon	/ɨ/
28.	/d/	bedydd	baptism	/ɛ/
29.		blodyn	flower	/ɔ/
30.		budr	dirty	/ɨ/
31.		dadl	argument	/a/

32		odl	rhyme	/ɔ/
33.		pedwar	four	/ɛ/
34.		rwdins	swede	/ʊ/
35.	/g/	dagr	dagger	/a/
36.		digon	enough	/ɪ/

Table A2: Word-final tokens

	Phoneme	**Token (Welsh)**	**English**	**Preceding vowel (based on auditory inspection and Mayr & Davies 2011)**
37.	/p/	ap	son of	/a/
38.		cip	glance	/ɪ/
39.		siop	shop	/ɔ/
40.		tap	tap	/a/
41.		twp	stupid	/ʊ/
42.	/t/	brat	apron	/a/
43.		cot	cot	/ɔ/
44.		cwt	hut	/ʊ/
45.		het	hat	/ɛ/
46.		sut	how	/ɨ/
47.	/k/	cic	kick	/ɪ/
48.		clec	smack	/ɛ/
49.		hac	to hack	/a/
50.		lwc	luck	/ʊ/
51.		toc	presently	/ɔ/
52.	/b/	gwlyb	wet	/ɨː/
53.	/d/	bod	to be	/oː/
54.		cwd	sack	/uː/
55.	/g/	cig	meat	/iː/

Table A3: Regression coefficients with *z* and *p* values for the model predicting the presence of breathiness.

Factor	**Level**	***ß***	**SE**	***z***	***p***
(Intercept)	**+1**	**1.95**	**0.64**	**3.06**	**< 0.01**
Vowel Category	Mid front	-0.82	0.57	-1.44	0.15
	High front	-1.00	0.53	-1.90	0.06
	Mid back	-0.75	0.56	-1.35	0.18
	High back	**-1.34**	**0.57**	**-2.37**	**0.02**
Place of Articulation	**Coronal**	**1.87**	**0.56**	**3.35**	**< 0.01**
	Velar	**1.87**	**0.57**	**3.26**	**< 0.01**
Consonant Series	Lenis	-1.51	0.81	-1.87	0.06
Sex	**Male**	**-1.42**	**0.61**	**-2.33**	**0.02**

Place of Articulation * Consonant Series	Coronal * Lenis	-0.62	0.97	-0.64	0.52
	Velar * Lenis	0.03	1.25	0.02	0.98

Table A4: Regression coefficients with *z* and *p* values for the model predicting the presence of pre-aspiration.

Factor	**Level**	***ß***	**SE**	***z***	***p***
(Intercept)	**+1**	**2.26**	**0.58**	**3.90**	**< 0.01**
Vowel Category	Mid front	-0.55	0.52	-1.05	0.29
	High front	**-1.25**	**0.47**	**-2.66**	**0.01**
	Mid back	**-1.13**	**0.50**	**-2.23**	**0.03**
	High back	**-1.84**	**0.50**	**-3.65**	**< 0.01**
Place of Articulation	**Coronal**	**1.54**	**0.48**	**3.21**	**< 0.01**
	Velar	**2.82**	**0.54**	**5.20**	**< 0.01**
Consonant Series	Lenis	-2.37	0.75	-3.17	< 0.01
Sex	**Male**	**-1.78**	**0.57**	**-3.14**	**< 0.01**
Place of Articulation * Consonant Series	Coronal * Lenis	-0.13	0.87	-0.15	0.88
	Velar * Lenis	-0.13	1.14	-0.12	0.91

Table A5: Regression coefficients with *t* and *p* values for the model predicting the duration of breathiness (ms).

Factor	**Level**	***ß***	**SE**	***t***	***p***
(Intercept)		**18.68**	**2.99**	**6.24**	**< 0.01**
Place of Articulation	**Coronal**	**5.74**	**2.40**	**2.40**	**0.02**
	Velar	4.47	2.62	1.71	0.09
Consonant Series	**Lenis**	**-6.56**	**2.12**	**-3.09**	**< 0.01**
Closure Duration (ms)		-0.02	0.01	-1.50	0.13

Table A6: Regression coefficients with *t* and *p* values for the model predicting the duration of pre-aspiration (ms).

Factor	**Level**	***ß***	**SE**	***t***	***p***
(Intercept)		**40.44**	**5.12**	**7.90**	**< 0.01**
Vowel Category	Mid front	-5.79	4.30	-1.35	0.18
	High front	**-13.01**	**4.07**	**-3.19**	**< 0.01**
	Mid back	-5.74	4.21	-1.36	0.18
	High back	**-15.28**	**4.39**	**-3.48**	**< 0.01**
	Coronal	3.05	4.29	0.71	0.48

Place of Articulation	**Velar**	**12.53**	**4.42**	**2.84**	**0.01**
Consonant Series	Lenis	**-16.61**	**6.50**	**-2.56**	**0.01**
Sex	**Male**	**-11.27**	**4.02**	**-2.81**	**0.01**
Home Language	**English**	**-8.99**	**4.02**	**-2.24**	**0.04**
Place of Articulation * Consonant Series	Coronal * Lenis	0.39	7.70	0.05	0.96
	Velar * Lenis	0.26	9.85	0.03	0.98

Table A7: Regression coefficients with *t* and *p* values for the model predicting the BP ZCR measurement of breathiness.

Factor	**Level**	***ß***	**SE**	***t***	***p***
(Intercept)		**1505.20**	**161.01**	**9.35**	**< 0.01**
Vowel Category	**Mid front**	**-264.75**	**117.60**	**-2.25**	**0.03**
	High front	**-366.36**	**111.69**	**-3.28**	**< 0.01**
	Mid back	**-473.01**	**115.09**	**-4.11**	**< 0.01**
	High back	**-721.05**	**120.45**	**-5.99**	**< 0.01**
Place of Articulation	**Coronal**	**325.79**	**102.56**	**3.18**	**< 0.01**
	Velar	**450.08**	**107.83**	**4.17**	**< 0.01**
Consonant Series	**Lenis**	**-318.79**	**90.21**	**-3.53**	**< 0.01**
Sex	**Male**	**-538.90**	**185.69**	**-2.90**	**0.01**

Table A8: Regression coefficients with *t* and *p* values for the model predicting the BP ZCR measurement of pre-aspiration.

Factor	**Level**	***ß***	**SE**	***t***	***p***
(Intercept)		**2001.56**	**235.54**	**8.50**	**< 0.01**
Vowel Category	Mid front	-141.65	202.66	-0.70	0.49
	High front	**-445.39**	**191.11**	**-2.33**	**0.02**
	Mid back	**-580.68**	**194.80**	**-2.98**	**< 0.01**
	High back	**-1064.69**	**203.92**	**-5.22**	**< 0.01**
Place of Articulation	**Coronal**	**678.33**	**173.00**	**3.92**	**< 0.01**
	Velar	**925.75**	**183.80**	**5.04**	**< 0.01**
Consonant Series	**Lenis**	**-1041.11**	**148.81**	**-7.00**	**< 0.01**
Sex	**Male**	**-760.96**	**241.72**	**-3.15**	**0.01**

References

Awbery, Gwenllian M. 1984. Phonotactic constraints in Welsh. In Martin J. Ball & Glyn E. Jones (eds.), *Welsh Phonology: Selected Readings*, 65–104. University of Wales Press: Cardiff.

Baayen, Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge, U.K.: Cambridge University Press.

Ball, Martin J. 1981. Descriptions of the vowels of North Wales: A Review. *Studia Celtica* 16/17, 191–208.

Ball, Martin J. 1984. Phonetics for phonology. In Martin J. Ball & Glyn E. Jones (eds.), *Welsh Phonology: Selected Readings*, 5–39. University of Wales Press: Cardiff.

Ball, Martin J. & Williams, Briony. 2001. *Welsh Phonetics*. Lampeter: Edwin Mellen Press.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2015. lme4: Linear Mixed-Effects Models using Eigen and S4 (Version 1.1-7). http://CRAN.R-project.org/package=lme4 (accessed 5 May 2016).

Boersma, Paul & David Weenink. 2015. Praat: Doing phonetics by computer (Version 6.0.28). http://www.praat.org (accessed 21 January 2015).

Clayards, Meghan & Thea Knowles. 2015. Prominence enhances voiceless-ness and not place distinction in English voiceless sibilants. In The Scottish Consortium for ICPhS 2015 (eds.) *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0934.pdf (accessed 18 May 2017).

Clayton, Ian. 2010. *On the Natural History of Preaspirated Stops*. Ph.D. dissertation, The University of North Carolina at Chapel Hill.

Clayton, Ian. 2017. Preaspiration in Hebrides English. *Journal of the International Phonetic Association*, doi: https://doi.org/10.1017/S002510031700007X, first published online 27 February 2017.

Cooper, Sarah E. 2015. *Intonation in Anglesey Welsh*. Ph.D. dissertation, Bangor University.

Coupland, Nikolas. 1980. Style-shifting in a Cardiff work-setting. *Language in Society* 9, 1–12.

DiCanio, Christian T. 2012. The phonetics of fortis and lenis consonants in Itunyoso Trique. *International Journal of American Linguistics* 78(2), 239–72.

Foulkes, Paul & Gerard Docherty. 1999. Derby and Newcastle: instrumental phonetics and variationist studies. In Paul Foulkes & Gerard Docherty (eds.), *Urban Voices: Accent Studies in the British Isles*, 47–71. Arnold: London.

Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Andersen, David Firth & Steve Taylor. 2016. Package 'effects': Effect Displays for Linear, Generalized Linear, and Other Methods. https://cran.r-project.org/web/packages/effects/effects.pdf (last accessed 18 May 2017).

Gordeeva, Olga & James Scobbie. 2007. Non-normative preaspirated voiceless fricatives in Scottish English: phonetic and phonological characteristics. *QMU Speech Science Research Centre Working Papers* 12. http://eresearch.qmu.ac.uk/154/1/WP12_Gordeeva_and_Scobbie_2007_v2.pdf (accessed 18 May 2017).

Gordeeva, Olga & James Scobbie. 2010. Preaspiration as a correlate of word-final voice in Scottish English fricatives. In Susanne Fuchs, Martine Toda & Marzena Żygis (eds.), *Turbulent Sounds: An Interdisciplinary Guide,* 167–207. Mouton de Gruyter: Berlin.

Grawunder, Sven, Sabine Asmus & Cormac Anderson. 2015. On the correlation of acoustic vowel and coda duration in Modern Welsh C(C)VC monosyllables. In The Scottish Consortium for ICPhS 2015 (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*. Glasgow: Glasgow University. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0574.pdf (accessed 2 April 2018)

Guion, Susan G. 2003. The vowel systems of Quichua-Spanish bilinguals: Age of acquisition effects on the mutual influence of first and second languages. *Phonetica, 60*, 98-128.

Hannahs, S.J, 2013. *The Phonology of Welsh*. Oxford: Oxford University Press.

Hejná, Michaela. 2015. *Pre-aspiration in Welsh English: A case study of Aberystwyth*. Ph.D. dissertation, University of Manchester.

Hejná, Michaela. 2016. *Pre-aspiration: manual on acoustic analyses 1.1*. https://ling.auf.net/lingbuzz/003184 (accessed 4 April 2017).

Hejná, Michaela & Wendell Kimper. Submitted. Pre-closure laryngeal properties as cues to the fortis-lenis plosive contrast in British English. In *Yearbook of the Poznań Linguistic Meeting 2017*.

Hejná, Michaela & Jane Scanlon. 2015. New laryngeal allophony in Manchester English. In The Scottish Consortium for ICPhS 2015 (eds.) *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0099.pdf (accessed 18 May 2017).

Helgason, Pétur. 2002. *Preaspiration in the Nordic Languages. Synchronic and Diachronic Aspects*. Ph.D. dissertation, Stockholm University.

Helgason, Pétur. 2003. Faroese preaspiration. In Maria J. Solé, Daniel Recasens & Joaquín Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*, 2517–20.

Helgason, Pétur & Catherine Ringen. 2008. Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36, 607–28.

Helgason, Pétur, Katrin Stölten & Olle Engstrand. 2003. Dialectal and sociophonetic aspects of preaspiration. In Maria J. Solé, Daniel Recasens & Joaquín Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*, 17–20.

Iosad, Pavel. Forthcoming. The [ATR]/Laryngeal connection and emergent features. In Bert Botma & Marc van Oostendorp (eds.), *Primitives of Phonological Structure*. Oxford University Press: Oxford.

Jones, Glyn E. 1984. The distinctive vowels and consonants of Welsh. In Martin J. Ball & Glyn E. Jones (eds.), *Welsh Phonology: Selected Readings*, 40–64. University of Wales Press: Cardiff.

Jones, Mark & Carmen Llamas. 2003. Fricated pre-aspirated /t/ in Middlesbrough English: an acoustic study. In Maria J. Solé, Daniel Recasens & Joaquín Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*, 655–8.

Kang, Yoonjung & Alexei Kochetov. 2010. An electroplatographic study of Korean cornoal obstruents. Presentation given at MOT. http://www.yoonjungkang.com/uploads/1/1/6/2/11625099/an_epg_study_of_korean_consonants.pdf (accessed 24 May 2017).

Karlsson, Anastasia & Jan-Olof Svantesson. 2011. Preaspiration in Mongolian dialects: acoustic properties of contrastive stops. *Proceedings of The 10th Seoul International Altaistic Conference*, 125–40.

Kettig, Thomas. 2015. *The BAD-LAD Split: A Phonetic Investigation*. MPhil dissertation, University of Cambridge.

Kuznetsova, Alexandra. 2015. lmerTest: Tests in Linear Mixed Effects Models (Version 2.0-25). http://cran.r-project.org/web/packages/lmerTest/index.html (accessed 5 May 2016).

Ladefoged, Peter & Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell

Laver, John. 1994. *The Phonetic Description of Voice Quality* Cambridge: Cambridge University Press.

MacWhinney, Brian. 2001. The Competition Model: the input, the context, and the brain. In Peter Robinson (ed.), *Cognition and Second Language Instruction*, 69–90. Cambridge: Cambridge University Press.

MacWhinney, Brian. 2012. The logic of the unified model. In Susan M. Gass & Alison Mackey (eds.), *The Routledge Handbook of Second Language Acquisition*, 211–27. London: Routledge.

Mayr, Robert. & Hannah Davies. 2011. A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh. *Journal of the International Phonetic Association* 41, 1–25.

Mayr, Robert, Jonathan Morris, Ineke Mennen & Daniel Williams. 2015. Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in Welsh and English. *International Journal of Bilingualism*, DOI: 10.1177/1367006915614921, first published online 26 November 2015.

McRobbie-Utasi, Zita. 1992. The role of preaspiration duration in the voicing contrast in Skolt Sámi. *Proceedings of the Second International Conference on Spoken Language Processing (ICSLP'92), Banff, Alberta, Canada, October 13-16 1992*, 1263–5.

Mees, Inger M, & Beverley Collins. 1999. Cardiff: A real-time study of glottalisation. In Paul Foulkes & Gerard Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 185–202. London: Arnold.

Morris, Jonathan. 2010. Phonetic variation in Northern Wales: preaspiration. In Miriam Meyerhoff, Chie Adachi, Agata Daleszynska & Ana Strycharz (eds.), *Proceedings of the Second Summer School of Sociolinguistics, The University of Edinburgh 14 - 20 June 2010*. Edinburgh: University of Edinburgh. http://orca.cf.ac.uk/64918/1/Jon.pdf (accessed 10 March 2017).

Morris, Jonathan. 2013. *Sociolinguistic variation and regional minority language bilingualism: An investigation of Welsh-English bilinguals in North Wales*. Ph.D. dissertation. Manchester: University of Manchester.

Morris, Jonathan. 2017. Phonetic variation in a long-term language contact situation: /l/-darkening in Welsh-English bilingual speech. *Journal of Sociolinguistics* 21(2), 183–207.

Nance, Claire & Jane Stuart-Smith. 2013. Pre-aspiration and post-aspiration in Scottish Gaelic stop consonants. *Journal of the International Phonetic Association* 43(2), 129–152.

Ní Chasaide, Ailbhe. 1985. *Preaspiration in Phonological Stop Contrasts*. Ph.D. dissertation, University College of North Wales.

Ní Chasaide, Ailbhe & Ó Dochartaigh, C. 1984. Some durational aspects of preaspiration. In J. W. Higgs & R. Thelwall (eds.), Topics in Linguistic Phonetics, in Honour of E.T. Uldall, 141–57. Belfast: New University of Ulster.

Ohala, John J. 2005. The marriage of phonetics and phonology. *Acoustical Science and Technology* 26(5), 418–422

Paulasto, Heli. 2006. *Welsh English Syntax: Contact and Variation* (University of Joensuu Publications in the Humanities 43). Joensuu: Joensuu University Press.

Piske, Thorsten, MacKay, Ian R.A. & Flege, James E. 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29, 191–215.

R Development Core Team. 2015. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.r-project.org (accessed 25 July 2015).

Ringen, Catherine & Wim van Dommelen. 2013. Quantity and laryngeal contrasts in Norwegian. *Journal of Phonetics* 41, 479–90.

RStudio. 2015. RStudio: Integrated development environment for R (Version 0.99.489). http://www.rstudio.com (accessed 25 July 2015).

Silverman, Daniel. 2003. On the rarity of pre-aspirated stops. *Journal of Linguistics* 39, 575–98.

Simonet, Miquel. 2010. Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *Journal of Phonetics* 38, 663-678.

Sloetjes, Han and Peter Wittenburg. 2008. Annotation by category - ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008). http://tla.mpi.nl/tools/tla-tools/elan/ (accessed 5 February 2014).

Spooner, Ciarán. 2016. *Provection in Ystalyfera Welsh*. M.A. dissertation, University of Edinburgh.

Stevens, Mary. 2010. How widespread is preaspiration in Italy? A preliminary acoustic phonetic overview. In Susanne Schötz & Gilbert Ambrazaitis (eds.), *Proceedings from Fonetik 2010, Lund, June 2-4, 2010*, 97–102. Lund: Lund University Centre for Languages and Literature.

Stevens, Mary. 2011. Consonant length in Italian: gemination, degemination and preaspiration. In Scott M. Alvord (ed.), *Selected Proceedings of the 5th Conference on*

Laboratory Approaches to Romance Phonology, 21–32. Somerville, MA: Cascadilla Proceedings Project.

Stevens, Mary & John Hajek. 2004a. How pervasive is preaspiration? Investigating sonorannt devoicing in Sienese Italian. In Steve Cassidy, Felicity Cox, Robert Mannell & Sallyanne Palethorpe (eds.), *Proceedings of the 10th Australasian International Conference on Speech Science and Technology*, 334–9. Sydney: Macquarie University.

Stevens, Mary & John Hajek. 2004b. Preaspiration in Sienese Italian and its interaction with stress in /VC:/ sequences. In Bernard Bel & Isabel Marlien (eds.), *Proceedings of Speech Prosody 2004,* 57–60. Nara, Japan: SProSIG.

Stevens, Mary & Ulrich Reubold. 2014. Pre-aspiration, quantity, and sound change. *Laboratory Phonology* 5(4), 455–488.

Thomas, Alan. R. 1994. English in Wales. In Robert Burchfield (ed.), *The Cambridge History of the English Language. Volume V – English in Britain and Overseas: Origins and Development*, 94–147. Cambridge: Cambridge University Press.

Thomas, Alan R. 1997. The Welshness of Welsh English: A survey paper. In Hildegard L.C. Tristram (ed.), *The Celtic Englishes*, 55–85. Heidelberg: Winter.

Thomas, Siân Elizabeth. 1988. A study of calediad in the upper Swansea Valley. In Martin J. Ball (ed.), The Use of Welsh, 33–6. Clevedon: Multilingual Matters.

Tronnier, Mechtild. 2002. Preaspiration in Southern Swedish dialects. *The Quarterly Progress and Status Report of the Department of Speech, Music and Hearing and Centre for Speech Technology* 44, 33–6.

Watson, Kevin. 2007. *The Phonetics and Phonology of Plosive Lenition in Liverpool English*. Ph.D. dissertation, University of Lancaster, Edge Hill College.

Wells, John C. 1982. *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

Welsh Language Commissioner. 2016. 2011 Census Results by Community. http://www.comisiynyddygymraeg.cymru/English/Policy,%20research%20and%20data/Census%20Data/Pages/2011CensusresultsbyCommunity.aspx (accessed 10 December 2016).

Footnotes

[1] In a thorough analysis of vowel and coda duration in C(C)VC monosyllables, Grawunder, Asmus & Anderson (2015) found clear durational differences between fortis and lenis plosives. Our findings indicate that there were no significant durational differences in either monosyllables or disyllables. It should be remembered, however, that we compared only *voiceless* fortis and lenis plosives in the analysis for the current study.