Whispering or Rumbling?:
A Survey of Welsh Rhotic Realisations

Joe Wellington (13344498)

Supervised by Cesko Voeten

BA Thesis Linguistics

Unviversity of Amsterdam

21ˢᵗ June 2025

Contents

Abstract

This study investigates the Welsh 'voiced' /r/ and 'voiceless' /r̥/ rhotic consonants among fluent speakers. It examines the maintenance of the distinction between the consonants in the context of language revitalisation and an influx of 'neo-speakers' from English-speaking backgrounds into the Welsh speaking community. The study investigates the relationship between the phonetic realisation of these sounds and the language background of the speakers. An acoustic analysis was conducted on the speech of 7 participants of various ages and language backgrounds from all across Wales. The participants were invited to take a questionnaire and 30 of their rhotic consonants were recorded in a word-reading task. The researcher measured the centre of gravity (CoG) for each segment and compared these between the two conditions. The proportion of each condition that was voiced was also measured and these two measurements were combined into a single 'acoustic distance' measurement. The analysis revealed that a majority of the participants (5 out of 7) maintained a statistically significant difference between the CoG values of each condition. The 'acoustic distance' measurement showed that the nature of this contrast was variable across participants, some participants had a large difference between their CoG values and small difference between the voicing proportions and vice versa. Contrary to previous literature, speakers from South Wales had a greater 'acoustic distance' between the conditions on average than speakers from North Wales. Though the study is limited by its small sample size, it suggests that the contrast between the two consonants is largely maintained in production and may be undergoing changes influenced by language background and age. The paper recommends further, larger scale investigations into the rhotics in order to better understand the changing sound-system of Welsh, which will be important in informing language planning policy for revitalising the language.

Keywords: Welsh, Rhotics, Language Revitalisation, Neo-Speakers, Phonetics

1 Introduction

1.1 The Revitalisation of Celtic Languages

The Welsh language: *Cymraeg,* spoken in Wales: *Cymru (/ˈkəm.ri/)* in the West of Great Britain, is a Celtic language. Though previously widespread, the Celtic languages have steadily declined and the surviving Insular Celtic languages, having outlasted their continental relatives (Stifter, 2008), have become minority languages in the vast majority of their range. Some of these languages, including those recently extinct like Manx (Ager, 2009) and Cornish (Ferdinand, 2013), entered periods of revival and revitalisation during the 20th and 21st centuries thanks to grass-roots organisations and governmental support. Previously shut out from the education system, the surviving Celtic languages are all used as languages of instruction in schools. Welsh, Breton and Scottish and Irish Gaelic all have schools teaching from a pre-school to a university level. Manx has pre-school and primary school instruction, and Cornish has pre-schools ("Mooinder Veggey", n.d.; Morton, 2013). The level of governmental support afforded varies significantly between the languages however. While Welsh speakers and Irish speakers in the Republic of Ireland have benefitted from equality in courts since 1993 and 1922 respectively ("Constitution of the Irish Free State (Saorstát Éireann) Act", 1922; "Welsh Language Act", 1993), Breton and Scottish Gaelic speakers are still required to speak the national language of their respective states in court if capable ("Gaelic Language (Scotland) Act", 2005). Breton speakers especially have faced strong resistance from the French state in achieving official status for their language, for example there have been several instances of Breton speakers being blocked from giving their children traditional Breton names due to them containing diacritics not present in French (James, 2023).

Among the extant Celtic languages, Welsh is considered the least threatened; the transmission rate of Welsh in households in which at least two adults could speak the language remained consistent at around 80% between the years 2011 and 2021, indicating the language will continue to persist between generations (Welsh Government, 2023). Additionally, Welsh enjoys both publicly funded and independent media organisations such as S4C and ITV Cymru, creating television programming and films which have even seen commercial success outside of Wales. The Welsh-speaking community also benefits from long-

running cultural institutions such as the *Eisteddfod*, a music and poetry festival which maintains the poetic traditions of the language and encourages their uptake by young Welsh-speakers (Sams, 2024). The Welsh government ambitiously aims to increase the number of speakers from approximately 538,300 (people aged three and above) (Welsh Government, 2022) to one million by the year 2050 by increasing the proportion of students in Welsh-medium education to 40%, promoting the usage of the language in day to day life and supporting the growth of Welsh-language culture (Welsh Government, 2018).

However, this revitalisation process faces challenges: as previously minoritised languages become languages of modern institutions, they often undergo rapid standardisation as institutional support favours a single speech variety, cluster of varieties or imposes a new 'neutral' variety (Kloss, 1967). This, combined with the integration of new speakers, can cause the speech of these "neo-speakers" to diverge significantly from traditional varieties, leading to rapid language change as the standardised language adopted by institutions can resemble a *koiné*: syncretising features of several different regional varieties resulting in a standardised language that is unfamiliar in at least some ways to traditional speakers. A particularly stark example of this is Breton: *Brezhoneg*, a Brythonic Celtic language spoken in Brittany in Northwestern France (Hewitt, 2016). Due to hostile national language policy and negative perception of the language, Breton experienced very low intergenerational transfer and declined rapidly in the latter half of the 20th century. While nationalist campaigns have had some success in raising Breton's status and accessibility, a significant gap has emerged between the declining traditional speech varieties and the 'Neo-Breton' taught in the *Diwan* immersion school system. Neo-Breton shows heavy French influence, particularly in phonology and syntax (Hornsby, 2005). This influence is so pronounced that Jones (1998) suggested that Neo-Breton and traditional Breton varieties are effectively separate languages, and traditional speakers of the language have commented that the neo-language feels 'artificial' and devoid of traditional figures of speech (Timm, 2001). This shows that despite an influx of language activists willing to learn a minority language, without careful language planning in consultation with heritage speakers the process may not be smooth and in extreme cases can result in two de-facto separate speech communities.

1.2 Variation and Change in Welsh

In Welsh, the language change driven by standardisation and incorporation of second language speakers isn't as pervasive or extreme as the case of Breton. Although effects are noticeable in very informal codes of Welsh, particularly in syntax and morphology such as in Figure 1 where the traditional VSO word order has been replaced with SVO such as in English. Throughout Wales an influx of new speakers educated in the Welsh language school system has driven the development of a new class of 'neo-speaker'. Coming from English-speaking backgrounds, the speech of these individuals is often characterised by eschewing many traditional words and constructions in favour of standardised language or calques and loanwords from English (see Figure 2).

Figure 1: Example of English Influenced Syntax

Standard Welsh

dw	i	wedi	bwyta	bara
am	i	PERF	eating	bread

'I have eaten bread'

Informal Welsh

fi	'di	bita	bara
i	PERF	eating	bread

'I have eaten bread'

Figure 2: Displacement of Regional Constructions

Informal South Welsh

Sai=n	gwylio=r	teledu
NEG.1SG.be=PROG	watching=DEF	television

'I'm not watching the television'

Standard Welsh

dw	i	ddim	yn	gwylio=r	teledu
am	i	NEG	PROG	watching=DEF	television

'I'm not watching the television'

One area of uncertainty concerns the status of the rhotic consonants in the Welsh language. Both Ball and Williams (2001) and Hannahs (2013) affirm that Modern Welsh has two rhotic consonants, /r̥/ and /r/, traditionally described as a voiceless and voiced alveolar trill respectively. The 'voiceless' variant occurs only at the beginning of words, having arisen from a historical sound change which caused rhotics to be devoiced word-initially. These two sounds do not occur in complementary distribution however, due to two phenomena: the adoption of loanwords from other languages beginning with /r/ (see Figure 3), and soft mutation (see Table 1). Soft mutation is a morphological process typical of the Brythonic Languages where certain consonants become more sonorous via voicing or spirantisation in the presence of a number of grammatical triggers. The literature differs on the extent to which they believe this contrast has been maintained in surviving varieties. Ball and Williams (2001) suggests the /r̥/ phoneme is often absent in many southern varieties, having merged with plain voiced /r/, whereas Hannahs (2013) states that the phonemes remain distinct. Ball and Williams (2001) concluded that further research on the precise phonetic character of these phonemes was needed. Following this, Jaworski and Asmus (2018) investigated the acoustic properties of complex trills, including Welsh /r̥/. Finding a range of realisation from those similar to the trilled 'canonical' realisations noted in pedagogical grammars of the language, to taps and flaps, and even finding fricativised realisations [ɹ^x - ɹ^ʃ] of the 'voiceless' phone, this shows that while the state of having two rhotics may persist among some Welsh speakers, it is not certain that the contrast between the two is a simple voicing distinction.

Figure 3: Examples of R-initial English Loanwords

rygbi
/ˈrəg.bi/
'rugby'

ras
/raːs/
race

radio
/ˈrad.jo/
'radio'

Table 1: The Soft Mutation

Radical	SM
/p/	/b/
/t/	/d/
/k/	/g/
/b/	/v/
/d/	/ð/
/g/	∅
/m/	/v/
/r̥/	/r/
/ɬ/	/l/

Given Welsh's longstanding and extensive contact with English, it's expected that some sound change might be motivated by this interaction. Ball and Williams (2001) provides evidence supporting the claim that the merger of /r̥/ and /r/ could result from language contact, noting it is more common among speakers of the Southern variety. This aligns with the greater phonetic similarity between English and Southern Welsh, including features like the loss of the /h/ phoneme (common in many UK English varieties), the fronting of the /ɨ/ phoneme to /i/, and the preservation of voiced /z/ in English loanwords where Northern Welsh would use /s/ (see Figure 4). This motivation for sound change is discussed by Bell (2013), who suggests that speakers undergoing language shift form a new dialect of the adopted language, this is exemplified by the disparate 'ethnic varieties' of English such as those spoken by African Americans in the USA and Maori people in New Zealand which display varying levels of influence from their substrate languages. This previously occurred in Wales, with the Welsh dialect of English exhibiting distinctive features shared with Welsh due to the Welsh-speaking population shifting to English (Paulasto et al., 2020) (see Figure 5). It could be argued that the current changes within Welsh, especially among neo-speakers, result from the reversal of this language shift, with previously English-speaking communities forming a distinct dialect group within the Welsh-speaking population (Hornsby, 2011).

Figure 4: Preservation of English /z/ in South Wales

South Welsh
zŵ
/zuː/
'zoo'

North Welsh
sŵ
/suː/
'zoo'

Figure 5: 'Emphatic There' in Welsh English[1]

Welsh English

there's	lovely	on	you
'how	lovely	for	you'

Welsh

dyna	hyfryd	i	ti
there	lovely	for	you

[1] Adapted from Paulasto et al. (2020) (p.108) (Researcher's translation)

1.3 The Present Study

Ultimately, this study's primary research question surveys the range of articulations for Welsh's two rhotic phonemes in speakers in Wales. It also seeks to explore the relationship between these realisations and the speakers' linguistic backgrounds and ages. It is hypothesised that participants from an English-language background will display less of a contrast between the two phones due to L1 interference. Given the lack of research into the phonetic quality of Welsh rhotics and the increasing relevance of investigating English L1 effects on new Welsh speakers, this is an opportune moment to research the language's current diversity. Such research is important to potentially inform language planning policy as the Welsh government aims to double the number of speakers by 2050 (Welsh Government, 2022). This study involved a series of interviews with fluent speakers of the language followed by a word reading task, focusing on an analysis of the rhotic consonants /r/ and /r̥/ in their speech. The study also utilized a short questionnaire to gather information on participants' gender, age, the area in which they learnt Welsh (or where

they grew up for speakers from Welsh-language households) and the language which was spoken in the home whilst they were being raised. Children were not surveyed as part of this study, as there is no reason to suggest that changes to the distribution of /r/ and /r̥/ are new developments specific to children, having been reported since at least 2001 (Ball and Williams).

2   Methodology

2.1   Participants

The target population for this research's participant pool comprised adult speakers of Welsh residing in and outside of Wales who self-identify as fluent, including speakers from both Welsh and English-speaking backgrounds. Despite attempts to recruit participants from both North and South Wales, the participant pool was biased towards speakers from North Wales, with 4 out of 7 analysed participants reporting coming from the north of the country. The participant pool was also made up mostly of speakers from a Welsh-speaking background with only 2 participants coming from an English-speaking or bilingual home environment.

These criteria were selected as the study aims to investigate the speech of members of the Welsh-speaking community, making fluency an important prerequisite for inclusion in the survey. Including speakers from both English and Welsh language backgrounds allowed the investigation of the impact of language background on the r-realisation of the specified phones. This is especially important as the Welsh government intends to drastically increase the Welsh speaking population, and it is highly likely that the majority of those new Welsh speakers will come from an English-speaking background. A snowball sampling strategy was used for participant recruitment; this approach was deemed most feasible as interviews took place using the online video conference sampling software Zoom due to the geographical distance between the participants and the researcher (Zoom Video Communications, Inc., 2025). The study recruited 12 participants in total, though due to time constraints and technical difficulties only 7 of the participants' reading task sections could be analysed. Segmentation of the audio was prioritised based on the independent variables of the participants in order to include a greater range of ages and language backgrounds in the dataset.

2.2 Procedure

Participants were recruited through the researcher's personal acquaintances, or through online communities of Welsh speakers and were invited to schedule an interview with the researcher. Prior to the start of the interview the participants were directed to fill out an online questionnaire (See Appendix A); the first question presented participants with an information brochure and sought their informed consent in order to continue. Subsequent questions then asked about participants' age, gender, language spoken at home and the part of Wales in which they learned Welsh. Once the questionnaire portion of the interview was complete, the researcher moved on to a semi-structured interview constisting of a number of friendly questions (See Appendix B). This section of the interview lasted around 10 minutes.

The following and final portion of the interview was a word reading task (See Appendix C): 15 words beginning with /r̥/ and 15 words with /r/ were presented to the participants. As the 'voiced' /r/ occurs infrequently in a word initial position, the two conditions used the same word conditioned by various prepositions and possessive pronouns that traditionally cause /r̥/ to be replaced with /r/ in a process called 'soft mutation'. This task was intended to account for the possibility that not enough realisations of the rhotic sounds were recorded during the semi-structured interview, though due to the previously mentioned technical difficulties this word reading task ended up forming the basis of the analysis.

2.3 Analysis

Following the interview, the data from the questionnaire and interview were immediately anonymised such that it could not be traced back to a participant's contact information. The questionnaire responses were linked to the participant's corresponding audio with an ID number immediately after the conclusion of the interviews. The participants' audio was then imported into Praat (Boersma & Weenink, 2025), and the silent portions were removed from the recording. The reading task section of the recording was isolated, and the

rhotic segments from the start of each word were segmented and extracted into separate audio files. These instances of rhotic segments were then analysed using Praat's built-in analysis tools; for each individual segment a centre of gravity (CoG) measurement was taken and recorded. This allowed for a measurement of the level of high frequency noise co-occurring with the segments. The presence of high frequency noise would indicate a pronunciation resembling that of the fricative-like realisations, [ɹˣ - ɹˢ] , as reported by Jaworski and Asmus (2018) (See Figure 6).

Figure 6: An instance of the spirantised 'voiceless' rhotic in *rhoi.*
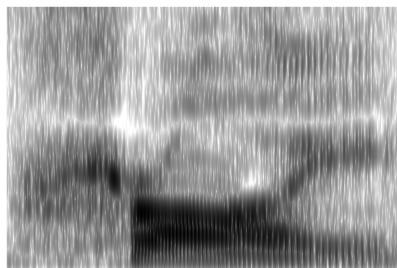
As the voicing of the segment is cited as a distinguishing feature of the two phonemes, the proportion of the segment which was voiced was also measured. Due to each extracted segment being short in duration, it was not possible to determine the voiced proportion of individual segments. In order to get around this limitation, all the segments of each condition for each participant were concatenated in order to create one 'voiced' and one 'voiceless' chain. These were then analysed using Praat's Analyse Periodicity function with the raw autocorrelation setting. The pitch floor and ceiling were set to an appropriate range of F0 frequencies so that only parts of the audio which exhibited a voicing bar would be marked by Praat as voiced. The appropriate values to use were informed by the gender of the participant as well as an examination of the participant's interview recording to observe their range. Praat then divided the concatenated chain of segments into a number of frames depending on its duration which were automatically marked as voiced or unvoiced. The voiced proportion of the condition was then expressed as the number of voiced frames

divided by the total number of frames.

The distinction between the rhotic phonemes seems to be multifaceted, with the contrast being realised variously as aspiration or high frequency noise and voicing. In order to capture this complex contrast, a single metric of each participant's average 'acoustic distance' between the two conditions was devised. The CoG values were standardised into z-scores, with each participant receiving their own mean and standard deviation. The voicing proportion data was then transformed using a logit transformation to convert the bounded proportions (from 0 to 1) into an unbound scale suitable for standardisation into z-scores. With the data on a comparable scale, the 'acoustic distance' between the two conditions was determined using the euclidean distance formula below (see Figure 7).

Figure 7: Formula for Acoustic Distance

$$\text{Acoustic Distance} = \sqrt{(\Delta\text{VoicingProportion\_z})^2 + (\Delta\text{CoG\_z})^2}$$

3   Results

Table 2: Participant Demographics

Participant ID	Age	Language Background	Regional Background	Gender
1	33	Welsh	South-East	Male
2	32	Welsh	South-East	Male
3	54	Welsh	North-West	Male
4	65	Welsh	Central	Female
7	18	Both	North-East	Male
8	54	Both	North-West	Female
10	23	English	North-West	Female

Table 3: Acoustic Data and Metrics

ID	Voiced CoG (Hz)	Voiceless CoG (Hz)	Voiced VProp.	Voiceless VProp	Acoustic Dist.
1	293*	396*	.967	.140	2.77
2	462	486	.796	.197	1.41
3	584**	1355**	.729	.0441	2.50
4	587**	742**	.929	.237	2.11
7	454*	682*	.359	.103	1.18
8	479*	1207*	.890	.180	2.03
10	512	549	.805	.590	0.623

* Indicates a statistically significant difference between the conditions. (p < .05)
** (p < .001)

3.1 Statistical Test Results

After the centre of gravity (CoG) datapoints were normalised, a paired Wilcoxon ranked sign test was performed for each participant in order to determine whether there was a statistically significant difference between the CoG values of each condition. This test was chosen as it could not be assumed that the original CoG data was normally distributed, making a non-parametric test more suitable. 5 of the 7 participants whose data was analysed had a significant difference between the two conditions, with two participants found to have a difference at a high level of significance (P3: p = .000183, P4: p = .000610). The p-values of individual tests ranged from P3: p = .000183 to P2: p = .762. A Wilcoxon ranked sign test at a group level showed a significant difference (p = .0156, r = -.9138) between the conditions for the analysed participants as a whole.

To investigate the relationship between the independent variables, age and language background, and the participants' z-scored CoG values, a linear regression model was employed. A model was run using the difference in CoG between the paired voiced and voiceless segments as the dependent variable; this resulted in no significant interaction, returning p = .166 for the variable of age, and p = .703 for language background.

The small sample size and the availability of only the total proportion of voicing for an entire condition, rather than for individual segments, constrain the use of inferential statistics. Consequently, the interpretation of the voiced proportion data and the 'acoustic distance' metric relies on descriptive statis-

tics. An analysis of the descriptive data reveals a notable variance between the voicing proportion and the Centre of Gravity (CoG) values across participants. For instance, as can be seen in Table 3, Participant 2 demonstrated a low, non-significant difference in CoG between conditions, yet a larger disparity in the voicing proportion. In contrast, Participant 7 showed a significant difference in CoG values between the two conditions, but a comparatively smaller difference in the voicing proportion than other participants.

The mean 'acoustic distance' of the group from a Welsh-speaking background was 2.20, while the mean 'acoustic distance' of the group of participants from English-speaking and bilingual backgrounds was 1.28, indicating that this group had a somewhat lower contrast in production between the two conditions. Somewhat unexpectedly, the mean 'acoustic distance' of participants from South and Central Wales (2.10) was greater than that of participants from North Wales (1.58). Although no significant influence of age on the difference between CoG values was found, it is notable that the participants with the two lowest 'acoustic distance' scores were also the youngest participants in the dataset. Language background was also not found to have a significant influence on the difference, although the participant (P7) with the second lowest acoustic distance score came from a bilingual household, while the participant with the lowest score came from an English-speaking household (P10), as can be seen in Tables 2 and 3.

3.2 Discussion

Due to a small participant pool, the survey's results cannot reliably be generalised to the wider Welsh-speaking population, however the findings provide some insight into trends amongst Welsh speakers. The group-level Wilcoxon test showed a significant difference (p = .0156), suggesting that the two Welsh rhotic phonemes remain distinct in production for many speakers as suggested by Hannahs (2013). While the nature of this distinction appears to be complex, the fact that a significant difference in CoG values could be found even in a participant with a relatively low difference in voicing proportion, such as participant 7, aligns with the findings of Jaworski and Asmus (2018) that the /r̥/ phoneme co-occurred with high frequency noise characteristic of frication or aspiration.

Despite the level of statistical significance, the descriptive statistics suggest a more nuanced picture that partially aligns with the study's initial hypothesis. The 'acoustic distance' metric designed to capture the combined contrast of CoG and voicing proportion reveals that participants who had English spoken at home while growing up exhibit a reduced contrast (mean distance of 1.28) than those who had grown up with predominantly Welsh spoken at home (mean distance of 2.20). Though this must be interpreted cautiously due to a low sample size, this suggests that Welsh speakers who acquired the language outside of the home or who became familiar with English at an earlier age may bring features from English into their Welsh. That the two youngest participants in the data set also had the lowest acoustic distance, lends credence to the idea that there may be an ongoing, age-graded change that warrants further research, even if such an effect was not detected in this study due to the small sample size.

One surprising finding was that the participants from South and Central Wales had a greater acoustic distance between the two segments than participants from North Wales. This seemed to contradict the findings of Ball and Williams (2001) that the two segments had merged in southern varieties of the language. Due to a small sample size, and that there were no participants both coming from South Wales and an English-speaking or bilingual background, it is difficult to judge how much credence should be lent to this finding.

The primary limitation of this study was the small sample size, which limited the statistical power of tests that were applied in analysis and resulted in the participant pool being biased. Ideally, the participant pool would have contained a more even spread of participants from different language backgrounds, ages and parts of Wales. Due to the snowball sampling method that was employed it was difficult to ensure the participant pool was representative of the Welsh-speaking population. One factor that likely contributed to the participant pool being biased towards speakers from North Wales is that Welsh is used as a community or professional language much more frequently in North Wales than in South Wales. This means that speakers in North Wales are more likely to be embedded in networks that allow for the invitation to be distributed more widely and picked up by a greater number of people. Whilst many participants from North Wales reported during the interview that they were colleagues at the same Welsh-speaking organisation, the relatively few

participants from South Wales were recruited through different networks and online communities. This may reflect the more 'atomised' state of the Welsh-speaking population in the south, as groups of speakers remain disconnected as a result of being expected to speak English as the default language, indeed all of the participants from North Wales who came from a Welsh-speaking background answered affirmatively when asked whether a lot of their friends speak Welsh, while this was not true of participants from South Wales.

Another factor that may have biased the participant pool towards speakers from Welsh-speaking backgrounds is a sense of language anxiety felt by prospective participants causing a self-selection bias. Especially in the South of the country where second language speakers have fewer opportunities to use the language, otherwise fluent Welsh speakers may not feel confident enough to volunteer to participate in a research project and therefore decline to respond to an invitation for an interview. This could explain how the participant pool became biased towards Welsh-speaking backgrounds and participants from North Wales despite the first round of invitations being sent to acquaintances of the researcher in South Wales.

As previously mentioned, technical problems limited the scope of the survey. The automatic speech recognition (ASR) model Whisper was planned to be used to assist in transcribing and segmenting the semi-structured interview section (Radford et al., 2023), although due to its inconsistency in transcribing fluent Welsh speech it was not able to produce usable output. This meant that only the word reading task was included in the final analysis, and several interviews that went ahead were not able to be segmented and analysed. This meant the difference between the realisations of the rhotic segments in spontaneous speech could not be compared with the realisations in the reading task, as well as reducing the number of data points available per participant. This resulted in a reduction in statistical power for modelling investigating the relationship between the realisations and the independent variables. The unreliability of Whisper for recognising and transcribing Welsh speech may be a product of Welsh's status as a minority language with less high quality training data available on which to train models and less incentive on the part of the developers to develop comprehensive support for Welsh. Whilst this issue may be solvable through the use of another ASR model created explicitly for Welsh, such as Trawsgrifwr (Canolfan Bedwyr, 2023). The use of Trawsgrifwr was

considered for this project, but was ultimately unable to be used as it requires the upload of audio files to a third party network, which would have breached ethical guidelines for the management of participant data.

Based on the preliminary findings and limitations of this survey, several avenues for further research into the quality of Welsh rhotics are apparent in order to clarify and possibly confirm some of the indicative findings of this project. A larger scale acoustic study is the most immediate path to producing results that could be more accurately generalised to the wider population. This study should be conducted in-person, ensuring a higher quality of audio recording than is possible over video conferencing software. In order to create a more representative participant pool, a quota-based sampling method should be employed to reflect the characteristics of the Welsh speaking population in terms of age and regional / language background. Most importantly, 'neo-speakers' of Welsh who grew up in English-speaking households should be targeted specifically for recruitment. Taking a proactive approach to including L2 speakers would be needed to counteract the aforementioned potential for selection bias, where speakers with lower confidence avoid participation. In addition to the CoG and voicing proportion measurements made in this survey, further research should also investigate the spectral skewedness of the rhotic segments in order to provide a clearer picture of the nature of the high frequency noise that can occur with the /r̥/ phoneme than can be provided with CoG measurements alone.

Another area of interest would be a longitudinal study to determine whether the reduced rhotic contrast tentatively observed in speakers with more English-influenced backgrounds is a permanent feature of L1 transfer or represents a temporary stage in their process of second language acquisition. This research would follow two cohorts of new Welsh speakers, one in South Wales and one in North Wales over an extended period of time. Acoustic measurements of their rhotic consonants /r̥/ and /r/ would be taken at regular intervals. Alongside this phonetic analysis, the study would also collect data on their language use, recording the amount of Welsh the participants speak on a daily or a weekly basis. If the participants' rhotic productions are shown to diverge over time, with the phonetic distance between the two conditions growing, it would suggest that the participants are acquiring the second rhotic phoneme over time. This would provide insight into whether the L1 influence that may

drive a lower contrast between the two phonemes is a permanent characteristic of 'neo-Welsh' or a developmental phase in their learning process.

4 Conclusion

This study set out to investigate the current phonetic status of the Welsh rhotic consonants /r/ and /r̥/ within the context of widespread and ambitious language revitalisation. Through acoustic analysis of participants from a range of different linguistic backgrounds, from those raised in Welsh-speaking home environments to newer Welsh speakers from English-speaking backgrounds, the research sought to determine whether the phonemic contrast is maintained or neutralised, especially among the growing population of new Welsh speakers.

The findings reveal a nuanced picture of the Welsh rhotic consonants. While a significant centre of gravity (CoG) distinction between /r/ and /r̥/ is retained for most speakers (5 of 7) in a formal reading task, the robustness of this contrast is highly variable, and the nature of participants' contrast between the two phonemes was not uniform, with some having a strong voicing contrast, while others had a strong CoG contrast. Although the trend was not statistically significant, the study provides an indication that there may be an influence of both age and language background on the degree of contrast exhibited between the rhotics.

The implications of these findings are significant for Welsh language planning. As the Welsh Government continues to pursue its ambitious goal of reaching one million speakers by 2050, it is inevitable that many of these speakers will be products of the Welsh-medium education system from non-Welsh speaking homes. The patterns observed in this study, with the youngest participants coming from English-speaking or bilingual home environments also having the lowest level of contrast between the phones, may then represent the future trajectory of the language. Understanding these shifts in pronunciation is important for informing teaching methods of the future and keeping instruction updated with the way the language is actually spoken, as well as for managing the relationship between the traditional heritage varieties of the language, and the emerging, standardised Welsh of an aspirationally bilingual country.

While constrained by a small sample size and reliance on a formal task, this study lays the groundwork for future investigation and confirms that a closer look at the evolving phonetics of Welsh is necessary. Larger scale and longitudinal studies will be necessary to gain more insight into the impact of absorbing a mostly monolingual English-speaking population into the Welsh-speaking community. Ultimately, continued research in this area is necessary for understanding and managing the consequences of large-scale language revitalisation, and for informing the policies that will shape the future of the language.

References

Ager, S. (2009). *A study of language death and revival with a particular focus on manz gaelic* [Doctoral dissertation, Bangor University (Linguistics)].

Ball, M. J., & Williams, B. (2001). *Welsh phonetics* (Vol. 17). The Edwin Mellen Press.

Bell, A. (2013). *The guidebook to sociolinguistics*. John Wiley & Sons.

Boersma, P., & Weenink, D. (2025). *Praat: Doing phonetics by computer* (Version 6.4.35). Retrieved June 15, 2025, from http://www.praat.org/

Canolfan Bedwyr. (2023). Datblygiadau diweddaraf yr uned technolegau iaith / latest developments from the language technologies unit [Presentation]. *Canolfan Bedwyr, Bangor University*.

Constitution of the irish free state (saorstát éireann) act [Irish Free State Act]. (1922). Retrieved June 21, 2025, from https://www.irishstatutebook.ie/eli/1922/act/1/enacted/en/print

Ferdinand, S. (2013). A brief history of the cornish language, its revival and its current status. *e-Keltoi: Journal of Interdisciplinary Celtic Studies, 2*(1), 6.

Gaelic language (scotland) act [Act of the Scottish Parliament]. (2005). Retrieved June 21, 2025, from https://www.legislation.gov.uk/asp/2005/7/contents

Hannahs, S. J. (2013). *The phonology of welsh*. OUP Oxford.

Hewitt. (2016). The problem of neo-speakers in language revitalization: The example of breton.

Hornsby, M. (2005). Néo-breton and questions of authenticity. *Estudios de Sociolingüística*, *6*(2), 191–218.

Hornsby, M. (2011). Colli iaith neu newid iaith? a critical sociolinguistic look at recent changes in the welsh language. *Poznań Linguistic Meeting*.

James, L. (2023, September). *French government face legal action after banning parents from giving their child a breton name*. Nation Cymru. Retrieved June 21, 2025, from https://nation.cymru/news/french-government-face-legal-action-after-banning-parents-from-giving-their-child-a-breton-name/

Jaworski, S., & Asmus, S. (2018). An acoustic study of welsh and slavonic rhotics. *Studia Celto-Slavica*.

Jones, M. C. (1998). *Language obsolescence and revitalization: Linguistic change in two sociolinguistically contrasting welsh communities*. Oxford University Press.

Kloss, H. (1967). 'abstand languages' and'ausbau languages'. *Anthropological linguistics*, *9*(7), 29–41.

Mooinder veggey. (n.d.). Mooinder Veggey. Retrieved June 21, 2025, from https://www.mooinjerveggey.org.im/#

Morton, K. (2013). *First cornish language pre-school to open*. Nursery World. Retrieved June 21, 2025, from https://www.nurseryworld.co.uk/content/news/first-cornish-language-pre-school-to-open/

Paulasto, H., Penhallurick, R., & Jones, B. (2020). *Welsh english* (Vol. 12). Walter de Gruyter GmbH & Co KG.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. Retrieved June 21, 2025, from https://github.com/openai/whisper

Sams, H. (2024). Finding their voices: The young-adult poets of the urdd national eisteddfod. *The International Journal of Young Adult Literature*, *5*(1), 1–23. https://doi.org/10.24877/IJYAL.138

Stifter, D. (2008). Old celtic languages: Gaulish. general information. *N/A*.

Timm, L. A. (2001). Transforming breton: A case study in multiply conflicting language ideologies. *Proceedings from the Eighth Annual Symposium about Language and Society—Austin*.

Welsh Government. (2018, December). *Cymraeg 2050: Welsh language strategy* (tech. rep.) (Strategy Document). Welsh Government. Retrieved June 21, 2025, from https://www.gov.wales/sites/default/files/publications/2018-12/cymraeg-2050-welsh-language-strategy.pdf

Welsh Government. (2022). *Welsh language in wales (census 2021)* (tech. rep.) (Statistics Document). Welsh Government. Retrieved June 21, 2025, from https://www.gov.wales/welsh-language-wales-census-2021-html

Welsh Government. (2023, June). *Welsh language composition of households in wales (census 2021)* (tech. rep.). Welsh Government. Retrieved June 21, 2025, from https://www.gov.wales/welsh-language-composition-households-wales-census-2021-html

Welsh language act [UK Public General Act]. (1993). Retrieved June 21, 2025, from https://www.legislation.gov.uk/ukpga/1993/38/contents?lang=en

Zoom Video Communications, Inc. (2025). *Zoom* (Version 5.17.11). https://zoom.us

Appendices

A   Questionnaire Questions

1. Please read the following consent forms and give your informed consent
   below:
   Darllenwch os gwelwch yn dda y ffurflenni cydsyniad canlynol ac rhowch eich
   cydsyniad gwybodus isod:
   I consent to / *Rwy'n cydsynio i*:

   Checkbox: Take part in this research / *Gymryd rhan yn yr ymchwil hwn*
   Audio recordings being made / *recordiadau sain yn cael eu gwneud*
   My personal details being stored for a period of 10 years / *fy manylion*
   personol yn cael eu storio am gyfnod o 10 Mlynedd

2. How old are you?
   Faint ydy dy oed di?

   Answer by sliding scale from 18 to 120

3. In which part of Wales did you learn Welsh? If you grew up speaking
   Welsh, where did you grow up?
   Ym mha ran o Gymru wyt ti wedi dysgu Cymraeg? Os wyt ti wedi tyfu lan yn
   siarad Cymraeg, ble wyt ti wedi tyfu lan?
   Multiple choice answer:
   North West Wales / *Gogledd-Orllewin Cymru*
   North East Wales / *Gogledd-Ddwyrain Cymru*
   South West Wales / *De-Orllewin Cymru*
   South East Wales / *De-Ddwyrain Cymru*
   Central Wales / *Canol Cymru*

4. What is your Gender?
   Beth ydy dy rywedd di?

   Multiple Choice Answer:
   Male / *Dyn*
   Female / *Merch*

Non-Binary / *Anneuaidd*
Prefer not to say / *Dw i ddim eisiau dweud*

5. What language was spoken at home while you were growing up?
Pa iaith cafodd ei siarad yn y tŷ pan ro't ti'n tyfu lan?

Multiple Choice Answer:
Predominantly Welsh / *Cymraeg yn Fwyaf*
Both Languages / *Y Ddwy iaith*
Predominantly English / *Saesneg yn Fwyaf*
Another language aside from English / *Iaith Arall ar Wahan Saesneg*

6. Do you consider yourself to be a fluent Welsh speaker?
Wyt ti'n siaradwr Cymraeg rhugl?

Binary Answer:
Yes / Ydw
No / Nac Ydw

B Interview Questions

- What do you like to do on the weekends? *Beth wyt ti'n hoffi gwneud ar y penwythnos?*

- What is your favourite sport?
Beth ydy dy hoff chwaraeon di?

- What is your favourite food?
Beth ydy dy hoff fwyd di?

- Where would you like to go next on holiday?
I ble hoffet ti fynd ar dy wyliau nesaf?

- Do you speak Welsh in your workplace?
Wyt ti'n siarad Cymraeg yn dy weithle di?

- Do many of your friends speak Welsh?
Ydy lot o dy ffrindiau yn siarad Cymraeg?

- What is your favourite animal and why?
  Beth ydy dy hoff anifail, a pham?

- Do you speak any other languages?
  Wyt ti'n siarad ieithoedd eraill?

C   Reading Task Stimuli

- Rhedeg - running

- Rhan - part

- Rhes - row

- Rhieni - parent

- Rhuban - ribbon

- Rhydd - free

- Rhosyn - rose

- Rhisgl - tree bark

- Rhisga - Risca (a town in South Wales)

- Rhagfarn - prejudice

- Rhyd - rust

- Rhwyd - net

- Rhyfedd - strange

- Rhad - cheap

- Rhodd - gift

- I Redeg - to run

- Heb Ryd - without rust

- Merch Ryfedd - a strange woman

- Heb Ruban - without a ribbon

- O Rwyd - from a net

- Ar Res - on a row

- Gan Rieni - from parents

- Ei Rodd - his gift

- Tra Rydd - whilst free

- O Risga - from Risca

- Dwy Ran - two parts

- Am Rad - how cheap

- Hen Rosyn - old rose

- Heb Risgl - without bark

- Heb Ragfarn - without prejudice