Automatic Speech Recognition Biases in Newcastle English: an Error Analysis

Dana Serditova[1,2], *Kevin Tang*[2], *Jochen Steffens*[3]

[1]English Department, University of Freiburg, Germany
[2]Faculty of Arts and Humanities, Heinrich Heine University Düsseldorf, Germany
[3]Institute of Sound and Vibration Engineering, University of Applied Sciences Düsseldorf, Germany

dana.serditova@anglistik.uni-freiburg.de, jochen.steffens@hs-duesseldorf.de,
kevin.tang@hhu.de

Abstract

Automatic Speech Recognition (ASR) systems struggle with regional dialects due to biased training which favours mainstream varieties. While previous research has identified racial, age, and gender biases in ASR, regional bias remains underexamined. This study investigates ASR performance on Newcastle English, a well-documented regional dialect known to be challenging for ASR. A two-stage analysis was conducted: first, a manual error analysis on a subsample identified key phonological, lexical, and morphosyntactic errors behind ASR misrecognitions; second, a case study focused on the systematic analysis of ASR recognition of the regional pronouns "yous" and "wor". Results show that ASR errors directly correlate with regional dialectal features, while social factors play a lesser role in ASR mismatches. We advocate for greater dialectal diversity in ASR training data and highlight the value of sociolinguistic analysis in diagnosing and addressing regional biases.

Index Terms: speech recognition, Newcastle English, ASR biases, error analysis, sociolinguistics

1. Introduction

One of the major challenges of Automatic Speech Recognition (ASR) that has been repeatedly highlighted in research is social bias. While racial, age, and gender biases are well-documented [1, 2, 3, 4], the investigations of regional biases remain limited [5, 6, 7]. This paper zooms in on ASR biases in Newcastle English, which is one of the most well-studied [8, 9, 10], yet one of the most demanding dialects for ASR systems [6].

1.1. ASR biases

ASR systems struggle with language variation since they rely on acoustic and language models trained on limited data and mainstream dialects. Recent studies have inspected various adaptation strategies such as speaker-specific models, multi-dialect training, and accent identification, but there is still a gap in research with regard to dialectal and sociolectal variation [5, 11, 12].

1.1.1. Racial bias

Research has demonstrated that existing ASR systems struggle with racial biases and exhibit higher Word Error Rate (WER) for marginalised communities. Koenecke et al. [1] found that Black speakers' WERs were nearly double those of White speakers across ASR systems from major companies. Liu et al. [3] reported disparities in WER across skin tone groups, and Wassink et al. [5] found significant variation in WERs amongst Native American, African American, European American, and ChicanX speakers. Beyond phonetic differences, Mar-

tin & Tang [13] showed that morphosyntactic features that are absent in Mainstream Englishes, such as "habitual be", a common AAE feature, increase WER in ASR.

1.1.2. Age and gender bias

Age and gender biases have been previously investigated and demonstrate a number of trends. Liu et al. [3] reported a consistent bias where male speakers receive higher WERs. The same gender-related result was demonstrated in Koenecke et al. [1], with male speakers obtaining higher WERs compared to females. Similarly, Feng et al.[1] found lower WERs for female speakers in a Dutch ASR system, with the largest difference observed amongst older adults and smallest amongst children. In terms of age, ASR performed best on teenage speech, and worst on children's speech, followed by older adults. However, the impact of age on ASR performance is not always consistent: age bias was not conclusive in Liu et al. [3], with WERs remaining stable across the sample.

1.1.3. Regional bias

With more obvious social biases taking centre stage in current speech technology research, regional and dialectal biases often remain overlooked. By dialectal bias, we refer to systematic ASR errors that can be directly attributed to and explained by regionally specific linguistic features, including phonological, morphosyntactic, and lexical variation characteristic of a given dialect or speech community.

Research has highlighted ASR struggles with language variation and underrepresented dialects [14]. Wassink et al. [5] found that sociophonetic variation contributes to ASR errors across ethnic dialects in the American Pacific Northwest. In the UK, ASR performance varies by region, with Southern British English speakers receiving the lowest WERs, while those from Northern England and Northern Ireland experience significantly higher error rates [6]. The study suggests that regional biases stem from underrepresentation of these dialects in ASR training data and further reinforce existing linguistic marginalisation. Similarly, Torgbi et al.[2] demonstrated that fine-tuning ASR on Scottish English improves recognition accuracy.

---

[1]S. Feng, O. Kudina, B. M. Halpern, and O. Scharenborg, *Quantifying bias in automatic speech recognition*, arXiv:2103.15122, 2021.

[2]M. Torgbi, A. Clayman, J. J. Speight, and H. T. Madabushi, *Adapting Whisper for Regional Dialects: Enhancing Public Services for Vulnerable Populations in the United Kingdom*, arXiv:2501.08502, 2025.

1.1.4. ASR and sociolinguistic research

Sociolinguistic insights can significantly improve ASR by revealing how dialectal variation drives recognition errors. Systematic error analysis can help identify problematic linguistic features that cause ASR errors. These insights can inform targeted data augmentation, fairness evaluations, and the development of more inclusive acoustic models that perform accurately across a greater range of speakers.

Wassink et al. [5] argue that current ASR systems struggle with dialectal variation because acoustic models are primarily trained on standard language varieties, while regional, ethnic, and social dialects are often unaccounted for. Markl [6] makes a similar observation and states that quantitative evaluation alone is not sufficient to understand the causes of ASR errors. Instead, the need for qualitative and context-sensitive analysis that takes linguistic variation and social factors into consideration is emphasised. Torgbi et al.[2] claim that WER is not a sufficient performance assessment metric and that ASR errors should be evaluated manually and qualitatively.

1.2. Newcastle English

Newcastle English is one of the most recognised ones in England [15] and is well-researched by sociolinguists [8, 9, 10]. As demonstrated by Markl [6], it is also one of the most challenging UK accents for ASR. This study focuses on the most salient phonological, morphosyntactic, and lexical features of Newcastle English that might be particularly problematic for ASR.

Within phonetic and phonological features, the most salient ones include FOOT/STRUT split absence (e.g., "fun" [fʊn]); lack of BATH retraction (e.g., "path" [paθ]); FACE and GOAT monophthongisation ([eː] and [oː]/[ɵ]); PRICE vowel ([ɛɪ] or [iː]); NEAR variation ([ɪə] or [ɛə]); rounded NURSE vowel [øː]; near glottalisation of /p, t, k/; 'g'-dropping [16, 9, 8, 17, 18, 19].

In terms of regional morphosyntactic features, the most prominent ones are unmarked plurals (e.g., "six month" and "three pound"); multiple negation; conflation of Past Tense and Past Participle forms (e.g., "they've broke it"); regional pronouns "yous" (2nd pl.), "wor" (meaning "our"), "us" in object position instead of "me"; regional verb forms such as "div" and "divn't" (regional forms of "do" and "don't"), "gan" ("go") [20, 16, 21, 9, 22].

2. Method

2.1. Dataset

To evaluate ASR performance on Newcastle English, we use the *Diachronic Electronic Corpus of Tyneside English* (DECTE), which is a representative corpus of dialect speech from the Tyneside area of North-East England. It comprises 72 hours of naturalistic, spontaneous speech from a diverse range of 160 speakers (excluding the interviewers) from the Newcastle area, distributed across 99 files [23]. The corpus is fully transcribed by human annotators, which makes the output highly accurate and an ideal source for evaluating ASR performance.

2.2. ASR system selection

With a multitude of ASR systems on the market today, choosing a system to test requires careful consideration. We based our selection on two primary criteria: first, the system had to represent state-of-the-art ASR technology, incorporating the latest advancements in deep learning and speech processing; second, the system had to be commercially deployed and/or accessible

to the average user, in order to ensure that our analysis reflects real-world application.

With these two factors in mind, we pre-tested the DECTE corpus on four ASR systems including both commercial solutions and more specialised open-source ASR: Google Cloud Speech-to-Text (`https://cloud.google.com/speech-to-text`), CrisperWhisper (an advanced variant of OpenAI's Whisper) [24], Deepgram Voice AI (`https://deepgram.com/`), and Rev AI (`https://rev.ai`). 10% of the dataset were selected to evaluate these four ASR systems. Google Cloud Speech-to-Text and Deepgram Voice AI were eliminated first since they significantly underperformed compared to CrisperWhisper and Rev AI, with WERs ranging from 40% to 60%. CrisperWhisper and Rev AI were tested against each other on a bigger subset of files. Rev AI consistently outperformed CrisperWhisper across all speaker demographics, with a lower WER of approximately 10% for most recordings. Across the entire dataset, Rev AI (set to English UK) achieved an average WER of 31.95%.

2.3. Data processing and analysis

Data analysis consisted of two stages.[3] First, a representative subsample of 16 DECTE files (32 speakers, with interviewers excluded from the analysis) transcribed by Rev AI was manually coded for errors. Errors were identified by comparing ASR output to human-checked transcriptions at word level. Two trained sociophoneticians (first author and an assistant) reviewed errors manually while listening to the audio. The analysis focused on dialectal bias and meaningful errors. By this we understand the types of errors that can be explained by dialectal variation and directly correlated with dialectal features in the region. The errors that were caused by noise, overlapping speech, or other indeterminate factors were excluded.

Errors were classified at two levels. First, by type: phonological, lexical, standardisation, morphosyntactic, and spelling. Standardisation errors involved replacing dialectal features with Standard Southern British English forms (e.g., "me life" → "my life", "telly" → "television"). The remaining four error types correspond to distinct linguistic domains. The second level involved finer-grained classification within each error type. For instance, within phonological errors, we specified whether the errors were connected to vowel quality, monophthongisation, glottalisation, etc. For morphosyntactic errors, we defined them based on the grammatical category or syntactic phenomena they affected (pronouns, verb paradigm, tense and aspect, etc.).

Overall, 1,076 meaningful errors were included in the analysis. For each error, aside from the aforementioned classification, information on the gender and age of the speaker was included. This allowed us to conduct sociolinguistic analyses of errors and examine how social factors predetermine ASR performance.

In the second stage, we expanded our analysis to the full dataset (160 speakers) and examined errors that could be attributed to the local pronouns "yous" and "wor", since pronoun errors constituted one third of all morphosyntactic errors (see Sec. 3.2.3). We implemented an automated extraction process using custom Python scripts, which performed string alignment between the ASR output and human-verified transcriptions and identified mismatches where the target pronouns were misrecognised. The scripts then extracted each instance along with its surrounding context. Demographic information about

---

[3]Data and code are available at: `https://doi.org/10.17605/OSF.IO/T8YXK`

the speaker was manually added for further sociolinguistic and statistical analysis. Given that noise has been shown to have impact on ASR inferences (e.g., [13]), we estimated the Signal to Noise ratio (SNR) of each audio file using the Waveform Amplitude Distribution Analysis (WADA-SNR) [25]. To assess the effects of speaker age and gender on ASR performance, we fitted mixed-effects logistic regression models using `glmer` function from the `lmerTest` package [26] in R (Version 4.4.1). Models were fitted separately for each pronoun with age and gender as fixed effects, speaker as a random effect, and noise level included as a covariate.

A key limitation is the absence of linguistic feature tagging, causing dialect patterns to be misclassified as errors. Incorporating feature tagging could provide a more accurate assessment of ASR performance on Newcastle English.

3. Results

The errors were analyzed with i) language-external variables, gender and age, and ii) language-internal variables consisting of error specifications and the relevant dialectal features.

3.1. Language-external variables

Figure 1 (left) shows the proportion of error types by gender. First, it is important to note that the male speakers in the dataset received higher error rates than their female counterparts (male speakers: 53.7% vs. female speakers: 46.3%, 2-sample test for equality of proportions: $\chi^2 = 11.6$, $p < 0.001$; overall distribution of errors: $\chi^2 = 5.948$, $p < 0.05$). As shown in Figure 1, there are some gender differences in the error distribution. Male speakers received more lexical errors, e.g., multiple misrecognitions of the local lexical items "nowt", "owt", "aye", as well as the more common items like "lass", "lads", "bloke", "quid" (males: 37.02% vs. females: 19.08%, $\chi^2(1) = 41.224$, $p < 0.001$).

Morphosyntactic, phonological, and spelling errors were in relatively equal proportions for both genders (morphosyntactic: 5.62% vs. 5.02%, $\chi^2(1) = 0.093$, $p > 0.05$; phonological: 56.43% vs. 49.48%, $\chi^2(1) = 4.901$, $p < 0.05$; spelling: 1.81% vs. 1.21%, $\chi^2(1) = 0.306$, $p > 0.05$). However, female speakers in the sample were associated with higher rates of standardisation errors (13.25% vs. 5.36% for males, $\chi^2(1) = 19.351$, $p < 0.001$). These errors include ASR 'correcting' the local pronouns (e.g., "me dad" → "my dad", "meself" → "myself", "wor" → "our" or "my", "hisself" → "himself"). There were also many instances of replacing "cause" → "because". Some local syntactic features were standardised as well, e.g., unmarked plurals ("four month" → "four months").

Figure 1 (right) shows the proportion of error types by age group. The most notable discrepancies appear in the lexical errors. Age groups 16-20 and 60+ received the highest number of lexical errors (19.3% and 25%, respectively), which visibly decrease for those between 41 and 60 (12%). As mentioned above, many of these were associated with local lexical items such as "nowt", "owt", or "wey". Toponyms also constituted a large proportion of these errors. There were repeated misrecognitions of "Sunderland", "North Shields", "Tyne", and "Toon". Even the term "Geordie", which refers both to the local dialect and its speakers, consistently suffered from misrecognitions.[4]

---

[4]The lexical item "aye" is omitted in Figure 1. Since two men in the subsample used it persistently and frequently and it was misrecognised every time, it could overwhelm the sample and skew the result.
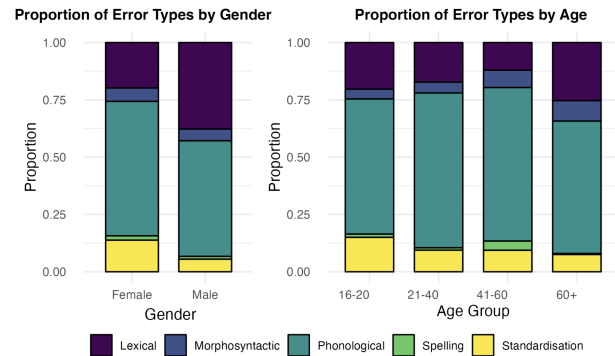

Figure 1: *Proportion of error types by gender and age.*

3.2. Language internal variables

Here, we examine language-internal factors behind ASR errors, analyzing phonological, morphosyntactic, and lexical errors in relation to local dialectal features. Examples of standardisation errors were given in Sec. 3.1. Spelling errors mostly consisted of spelling differences between British English and American English, where British spellings were substituted with an American alternative, and are not numerous.

3.2.1. Lexical errors

Unsurprisingly, lexical errors offer the most range and are particularly difficult to analyse quantitatively. We do not delve into details of these errors in the framework of this paper and offer only an overview. As expected, local lexical items presented the most difficulties for ASR. Amongst those, "nowt" and "owt" (meaning "nothing" and "anything") frequently posed challenges for transcription and were often mistranscribed as "now" or "no" and "out". Other local lexical features also led to errors, including "lush" (positive adjective), "geet" (emphatic "very"), "canny" (versatile, typically positive), "mint" ("excellent"), and "wey" (another versatile expression that can mean "well", "of course", or used to express surprise).

3.2.2. Phonological errors

Phonological errors related to glottalisation, monophthongisation, and vowel quality were the most prominent. Near glottalisation of /t/ was the most challenging feature for ASR, causing nearly 23% of phonological errors. These errors directly correlate with a well-documented feature of Newcastle English, where medial /p/, /t/, /k/ undergo glottalisation [17, 16, 9, 8]. Examples of such errors are "skipping" [skɪp͡ʔɪn] → "skiing" [skiːɪŋ], "patent" [peɪt͡ʔənt] → "paint" [peɪnt], "speaking up" [spiːk͡ʔɪn ʊp] → "speeding up" [spiːdɪŋ ʊp].

Monophthongisation of FACE [eː] accounted for 17.5% of errors. This trend, led by young, middle-class females in Newcastle [19, 8] may challenge ASR systems trained on more conservative speech. This feature caused errors such as "tray" [tɹeː] → "tree" [tɹiː], "whale" [weːl] → "wheel" [wiːl]. The third most common error was 'g'-dropping [ɪn] [9, 16], with over 11% of errors caused by this feature. Examples of errors include "drawing" [dɹɔːɪn] → "drawn" [dɹɔːn] and "wedding" [wɛdɪn] → "wind" [wɪn(d)].

The remaining major error types are GOAT-monophthongisation (8.4%), the glottalisation of /k/ and

/p/ (6% and 4.2%, respectively), and glottal stops (4%). Vowel quality was another major source of ASR errors (19.6%), with FLEECE, NURSE, START vowels, as well as NEAR variation causing the most misrecognitions.

3.2.3. Morphosyntactic errors

Local pronoun usage (38.6%), verb paradigm (42.1%), and tense and aspect (15.8%) were most challenging morphosyntactic errors for ASR. As discussed in Sec. 1.2, Newcastle English is characterised by non-standard pronoun usage. Amongst these, the pronouns "yous" (2nd pers. pl.), "wor" (meaning "our"), and 1st pers. sing. possessive "me" caused a third of all morphosyntactic errors (29.9%). Verb paradigm, consisting of non-standard variation such as "div" and "divnt", first-person singular verb -s, and "gan", caused another third of errors in this category (35.1%). Often, there was no output for these items (22% of cases) or a phonetically similar item was used (e.g., "didn't", "gun", "gone") (13% of cases). Finally, Past Perfect was persistently mistranscribed as Present Perfect by ASR, which can be explained by frequency bias (7% of cases).

3.3. ASR performance on local pronouns: yous and wor

After conducting a qualitative analysis of all errors in the subsample and identifying recurring error patterns, we expanded our analysis to the full dataset. Here, we present a systematic analysis of errors affecting local pronouns "yous" and "wor", for which we implemented an automated extraction process to identify and quantify misrecognitions at scale.

These pronouns were selected for further analysis because they represent a salient morphosyntactic feature of Newcastle English that cannot be found in Standard Southern British English. Furthermore, these pronouns are frequently misrecognised according to our manual error analysis in Sec. 3.2.3.

The results of the regression for the two pronouns are presented below. The models tested whether ASR produced the standardised forms ("yous" → "you", "wor" → "our") or an incorrect output as the dependent variable. For "yous", the model did not find a significant effect of age or gender ($ps > 0.05$). However, the intercept was significant ($\hat{\beta} = -1.22$, $p < 0.01$), which indicates overall low recognition accuracy for this pronoun. For "wor", the model revealed a significant effect of age, with speakers aged 21–40 significantly more likely to be correctly recognised compared to the reference group 16-20 ($\hat{\beta} = 1.79$, $p < 0.05$). Additionally, higher noise levels were associated with significantly lower ASR accuracy ($\hat{\beta} = -0.13$, $p < 0.05$) (Table 1).

Table 1: *Regression summary for ASR Recognition of "wor"*

Fixed Effect	Estimate	Std. Error	p-value
Intercept	-0.3077	0.9156	0.7368
Age: 21–40	1.7853	0.8735	0.0410 *
Age: 41–60	1.4596	0.8310	0.0790
Age: 60+	2.4948	1.7916	0.1638
Gender: Male	-1.6679	1.5082	0.2688
Noise (dB)	-0.1279	0.0503	0.0111 *

Random effect: Speaker ($\sigma^2 = 0$).

4. Discussion

Overall, it is clear that the lexical and phonological features are the primary challenge for ASR and vary the most across genders and ages. The proportion of phonological errors is highest across all speaker demographics. Within these, errors that correlated with monophthongisation, glottalisation, and overall vowel quality were the most common. These errors should be studied systematically in future research. Methodologically, automating the annotation of dialectal features is a crucial step in examining these errors at scale [27, 28].

The higher proportion of lexical errors for men likely reflects their greater use of non-standard forms, which are associated with traditional masculinity and local identity. This is linked to covert prestige, where non-standard language forms carry social value within specific communities [29, 30]. By contrast, research in sociolinguistics has consistently highlighted that women favour standard forms [29, 30, 31, 32], and ASR models trained on mainstream data may reinforce these biases.

The variation in error proportions, particularly in the lexical errors, can be explained by age grading — a sociolinguistic process where speech patterns shift across life stages. Younger speakers frequently use more non-standard forms as linguistic innovators, while working adults tend to adopt standard forms due to professional and social pressures, and the influence of education. In later life, as these pressures decrease, speakers may return to vernacular features acquired in youth. Studies on Newcastle English confirm this pattern for 'g'-dropping, the GOAT vowel, and possessive "me", and conclude that vernacular use peaks in adolescence, declines during adulthood, and increases again in retirement [18, 33, 34, 22].

The statistical analysis of ASR errors, particularly for the pronoun "wor", supports this claim. The results indicate that ASR struggles the least with this feature amongst working adults, while younger and older speakers receive more misrecognitions. These findings show that age grading interacts with ASR biases and that sociolinguistic patterns play a bigger role in ASR errors. Therefore, attributing these errors solely to model limitations is an oversimplification.

It is also evident from our analyses that ASR errors are driven more by language-internal, dialect-specific factors than by language-external, social factors, despite some variation across speaker demographics. Thus, future research should address dialectal errors and acknowledge that social variables do not necessarily explain the full scope of ASR misrecognitions.

5. Conclusion

The paper investigated ASR performance on Newcastle English and demonstrated that ASR errors can be directly attributed to local phonological, lexical, and morphosyntactic features. Through a manual error analysis that provided an overview of all error types, and a case study that looked into pronoun errors systematically, we established that language-internal, dialectal errors play a more substantial role in ASR mismatches than social factors. The results highlight the need for dialectal diversity in ASR training data [35] and the value of sociolinguistic insights in error assessment.

6. Acknowledgements

7. References

[1] A. Koenecke, A. Nam, E. Lake, J. Nudell, M. Quartey, Z. Mengesha, C. Toups, J. R. Rickford, D. Jurafsky, and S. Goel, "Racial disparities in Automated Speech Recognition," *Proceedings of the National Academy of Sciences*, vol. 117, no. 14, pp. 7684–7689, 2020.

[2] S. Feng, B. M. Halpern, O. Kudina, and O. Scharenborg, "Towards inclusive Automatic Speech Recognition," *Computer Speech & Language*, vol. 84, p. 101567, 2024.

[3] C. Liu, M. Picheny, L. Sarı, P. Chitkara, A. Xiao, X. Zhang, M. Chou, A. Alvarado, C. Hazirbas, and Y. Saraf, "Towards measuring fairness in speech recognition: Casual conversations dataset transcriptions," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 6162–6166.

[4] G. Zellou, M. Cohn, and B. Ferenc Segedin, "Age-and gender-related differences in speech alignment toward humans and voice-AI," *Frontiers in Communication*, vol. 5, p. 600361, 2021.

[5] A. B. Wassink, C. Gansen, and I. Bartholomew, "Uneven success: Automatic Speech Recognition and ethnicity-related dialects," *Speech Communication*, vol. 140, pp. 50–70, 2022.

[6] N. Markl, "Language variation and algorithmic bias: understanding algorithmic bias in British English Automatic Speech Recognition," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 521–534.

[7] L.-F. Lai, J. G. van Hell, and J. Lipski, "Dialect bias in Automatic Speech Recognition: Analysis of Appalachian English," *American Speech: A Quarterly of Linguistic Usage*, pp. 1–17, 2024.

[8] A. Mearns, "Tyneside," in *Researching Northern English*. John Benjamins Publishing Company, 2015, pp. 161–182.

[9] A. Hughes, P. Trudgill, and D. Watt, *English Accents and Dialects: An introduction to social and regional varieties of English in the British Isles*. Routledge, 2013.

[10] E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, and C. Upton, *A Handbook of Varieties of English: A multimedia reference tool two volumes plus CD-ROM*. De Gruyter Mouton, 2004.

[11] M. K. Ngueajio and G. Washington, "Hey ASR system! Why aren't you more inclusive? Automatic Speech Recognition systems' bias and proposed bias mitigation techniques. A literature review," in *International Conference on Human-Computer Interaction*. Springer, 2022, pp. 421–440.

[12] R. Tatman and C. Kasten, "Effects of talker dialect, gender & race on accuracy of Bing speech and YouTube automatic captions," in *Interspeech*, 2017, pp. 934–938.

[13] J. L. Martin and K. Tang, "Understanding racial disparities in Automatic Speech Recognition: The case of habitual "be"." in *Interspeech*, 2020, pp. 626–630.

[14] R. Sanabria, N. Bogoychev, N. Markl, A. Carmantini, O. Klejch, and P. Bell, "The Edinburgh international accents of English corpus: Towards the democratization of English ASR," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[15] C. Montgomery, "The effect of proximity in perceptual dialectology," *Journal of Sociolinguistics*, vol. 16, no. 5, pp. 638–668, 2012.

[16] J. Beal, "English dialects in the North of England: Phonology," *A Handbook of Varieties of English*, vol. 1, pp. 113–133, 2004.

[17] G. Docherty and P. Foulkes, "Sociophonetic variation in 'glottals' in Newcastle English," in *Proceedings of the 14th International Congress of Phonetic Sciences*. University of California, Berkeley, 1999, pp. 1037–1040.

[18] J. Grama, J. Mechler, L. Bauernfeind, M. E. Eiswirth, and I. Buchstaller, "Post-educator relaxation in the U-shaped curve: Evidence from a panel study of Tyneside (ing)," *Language Variation and Change*, vol. 35, no. 3, pp. 325–350, 2023.

[19] D. Watt and L. Milroy, "Patterns of variation and change in three Newcastle vowels: Is this dialect levelling?" in *Urban Voices*. Routledge, 2014, pp. 25–46.

[20] J. Beal, "The grammar of Tyneside and Northumbrian English," in *Real English*. Routledge, 1993, pp. 187–213.

[21] M. Pearce, "Folk accounts of dialect differences in Tyne and Wear," *Dialectologia et Geolinguistica*, vol. 20, no. 1, pp. 5–25, 2012.

[22] A.-M. Moelders, "Navigating the vernacular across the lifespan: A panel study of the phonetic realisation of the first-person singular possessive," *English Language & Linguistics*, pp. 1–27, 2025.

[23] K. P. Corrigan, I. Buchstaller, A. Mearns, and H. Moisl, "The Diachronic Electronic Corpus of Tyneside English," *Online at https://research.ncl.ac.uk/decte/index.htm (Accessed April 2, 2024)*, 2012.

[24] M. Zusag, L. Wagner, and B. Thallinger, "Crisperwhisper: Accurate timestamps on verbatim speech transcriptions," in *Interspeech 2024*, 2024, pp. 1265–1269.

[25] C. Kim and R. M. Stern, "Robust signal-to-noise ratio estimation based on waveform amplitude distribution analysis," in *Ninth Annual Conference of the International Speech Communication Association*, 2008.

[26] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[27] D. Villarreal, L. Clark, J. Hay, and K. Watson, "From categories to gradience: Auto-coding sociophonetic variation with random forests," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 11, no. 1, p. 6, 2020.

[28] T. Kendall, C. Vaughn, C. Farrington, K. Gunter, J. McLean, C. Tacata, and S. Arnson, "Considering performance in the automated and manual coding of sociolinguistic variables: Lessons from variable (ing)," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

[29] P. Trudgill, "Sex, covert prestige and linguistic change in the urban British English of Norwich," *Language in Society*, vol. 1, no. 2, pp. 179–195, 1972.

[30] W. Labov, "The social motivation of a sound change," *Word*, vol. 19, no. 3, pp. 273–309, 1963.

[31] L. Milroy, *Language and Social Networks*. Blackwell, 1980.

[32] P. Nichols, "Linguistic options and choices for Black women in the rural South," *Language, Gender and Society*, pp. 54–68, 1983.

[33] L. Bauernfeind, C. Ahrens, J. Grama, R. Skarnitzl, and J. Volín, "Change across the lifespan in goat: Evidence from a panel study of Tyneside English," in *Proceedings of the 20th International Congress of Phonetic Science. Guarant International*, 2023, pp. 2064–2068.

[34] J. Grama, M. E. Eiswirth, I. Buchstaller, R. Skarnitzl, and J. Volín, "Tracking creak from early to late adulthood: A panel study from the North East of England," in *Proceedings of the 20th International Congress of Phonetic Science. Guarant International*, 2023, pp. 2049–2053.

[35] J. Grieve, S. Bartl, M. Fuoli, J. Grafmiller, W. Huang, A. Jawerbaum, A. Murakami, M. Perlman, D. Roemling, and B. Winter, "The sociolinguistic foundations of language modeling," *Frontiers in Artificial Intelligence*, vol. 7, 2025.

---

[5] https://credit.niso.org/