Accepted version. To appear in Linguistics Vanguard.
Follow up project info: https://congzhang-linguist.github.io/project/gamification/

Journal Article

Yanyu Li, Damar Hoogland, Daniel Duncan, and Cong Zhang*

Gamification in Phonetic Perception Data Collection: Accent Identification in North East England

Abstract: Gamification has been increasingly adopted across a range of domains, including academic research, thanks to its inclusive and scalable nature. Gamified tasks involve the incorporation of game elements such as points, levels, and rewards within non-game contexts to enhance participant' engagement and motivation. Despite its online accessibility and motivating nature, however, there is no agreed protocol regarding which gamified elements to include in what type of tasks, especially in linguistic research which is mostly text-based. Rather than proposing a one-size-fits-all solution, this paper demonstrates the viability of audio-based gamified tasks, testing whether listeners from the North East of England, a region in which many residents believe there are fine-grained dialectal distinctions, can accurately distinguish speakers by location. In comparison to traditional identification tasks, our tasks were gamified through the inclusion of elements such as a points system, a post-task leaderboard, and rewards for high-scoring participants. Results indicate that within-North East speaker identification was a more difficult task for listeners than folk belief would indicate. We suggest that despite its limitations, gamification successfully replaced traditional per-participant incentives as a data collection strategy in a phonetic perception study.

Keywords: gamification, data collection, citizen science, sociophonetics, English accents

Yanyu Li, Daniel Duncan, Cong Zhang, Newcastle University, United Kingdom, Emails: y.li218@newcastle.ac.uk; Daniel.Duncan@newcastle.ac.uk; Cong.Zhang@newcastle.ac.uk
Damar Hoogland, Jožef Stefan Institute, Slovenia, Email: damar.hoogland@ijs.si

1 Background

The rise of digital technologies and growing public preference for interactive and entertainment-based experiences have contributed to the increasing adoption of gamification in research areas such as economics, health, education, and marketing (Landers et al. 2018, Bozkurt & Durak 2018, Long et al. 2023, Shortt et al. 2023, Seaborn & Fels 2015). When implemented through standardised, scalable online gamified tasks, gamification can support reproducibility and higher statistical power by enabling consistent replication, and can support generalisability by facilitating broader recruitment, more diverse samples, and cross-context replication (Long et al. 2023, Hamari, Koivisto & Sarsa 2014, Shahzad et al. 2023, Hartshorne, Tenenbaum & Pinker 2018). For example, over four million people have participated in *The Music Lab* (Cho et al. 2025, Bertolo et al. 2025), which investigates perception and creation of music through large-scale gamified online experiments.

In linguistics, games have also been used for data collection, although most have been text-based rather than speech-based (e.g., Genovese et al. 2024, Leemann, Derungs & Elspaß 2019, Kim, Kogan & Zhang 2023). A small number of gamified designs have been used to collect speech production data, typically for large-scale crowdsourcing (e.g., Leemann et al. 2016, Leemann, Kolly & Britain 2018). While gamification can support large-scale data collection, its primary value is not necessarily maximising participant numbers; it can also improve participant experience in studies with relatively small samples.

Here, we aim to test whether gamification is a useful data collection method in phonetic research. Data collection in phonetics and phonology often requires participants to engage in repetitive listening or speaking tasks, which can be time-consuming and monotonous. Introducing gamification into this process offers a positive interference by enhancing participant motivation and encouraging more naturalistic responses. This approach can lead to larger datasets and more spontaneous speech, both of which are essential for advancing phonetic research. The current study implements a gamified accent identification task as a proof-of-concept test case to demonstrate how gamification methods can be used in phonetic data collection (RQ1). For this test case, we examine whether listeners from North-East England can distinguish different local North-East English accents accurately (RQ2) and if the selected phonetic features can predict the identification accuracy (RQ3).

1.1 Gamification: definition and elements

Game-related methods can be broadly categorised into two types (Deterding et al. 2011, Kim, Kogan & Zhang 2023, Kogan 2025): game-based methods and gamification methods. Game-based methods incorporate existing games. This is relatively common in phonetic research, where researchers often use structured activities such as map tasks (Thompson et al. 1993) or off-the-shelf games (e.g., Arvaniti et al. 2022) to elicit spontaneous speech . In contrast, gamification methods, not aiming for a full-fledged game, are usually defined as the incorporation of game elements — such as points, levels, competition and rewards — in non-game contexts to enhance engagement and motivation. These gamified elements tap into participants' intrinsic motivation: the internal desire to engage in an activity out of personal interest and satisfaction, rather than external pressure (Ryan & Deci 2000a,b, Bozkurt & Durak 2018). The aim of gamification is to improve participant experience, enhance engagement and potentially increase the scalability of data collection, which is intrinsically different from crowdsourcing data collection (e.g., Genovese et al. 2024, Eryiğit, Şentaş & Monti 2023, Hilton 2021).

Self-Determination Theory (SDT) (Ryan & Deci 2000b) posits that intrinsic motivation is driven by *autonomy* (feeling in control), *competence* (feeling capable), and *relatedness* (feeling connected). Gamified elements based on SDT are, for example, a progress bar that breaks down tasks into manageable steps, giving users the perception of autonomy (Kim 2015, Zichermann & Cunningham 2011, Ryan & Deci 2000a,b). Points and leaderboards serve as indicators of achievement by ranking players according to their performance, which can enhance participants' sense of competence. They may also create a sense of *relatedness* by providing recognition within a competitive context. For example, appearing on a leaderboard can enhance a player's visibility and social standing within a group.

Therefore, gamification may help make cognitively demanding tasks feel more approachable and less monotonous. This can be particularly relevant for phonetic research, where data collection is often long, repetitive, and costly. Incorporating gamification could also support citizen science in phonetics by encouraging broader participation in the research process. In addition, gamification can act as a mediator for raising awareness of socially relevant issues and provide a way of connecting academic research with public engagement. Thanks to their interactive and playful nature, gamified approaches may also attract the general public more readily, for example through knowledge exchange events.

1.2 Accent identification: an overview

Speakers draw inferences about their interlocutor(s) based upon indexical relations between linguistic features and social categories (Eckert 2008). One such social category is one's region of origin. However, regional origin intersects with other social categories such as race, class, and ethnicity, shaping how people perceive and evaluate dialect boundaries. This intersection produces perceptions of prestige vs. non-standard dialects, affecting social solidarity and identity (Coupland & Bishop 2007).

Even if listeners can reliably identify varieties that they are familiar with, it is unclear how fine-grained this identification can be. Residents of a metropolitan area often claim both that the local language variety differs between towns and neighbourhoods, and that they can recognise such differences. One well-known instance of this is that New Yorkers often claim that the five boroughs of New York City have distinct accents. Crucially, these folk claims involve categorical differences in kind, rather than differences in frequency of usage; for example, it is claimed that Queens is "nasal" to the exclusion of other boroughs (Becker & Newlin-Łukowicz 2018). Such claims are usually at odds with variationist sociolinguistic research, which finds that speakers from all over a city are members of the same speech community and therefore share both evaluative norms and a "uniformity of abstract patterns of variation" (Labov 1972: 121). This would include shared access to the same variants of a given variable, though of course speakers will vary systematically in usage rates. That said, social categories that influence patterns of linguistic variation such as social class are very often spatialised. As such, class-based variation in a single regional variety may present as regional variation within a city or larger metropolitan area (Duncan 2019). Given this spatialisation of class- and other social category-based variation, it is possible that local listeners attuned to patterns of linguistic variation in their speech community may be able to identify which neighbourhood a speaker is from by relying on their local geographical knowledge in conjunction with their knowledge of how a linguistic feature indexes a social category. However, explicit tests of this, such as that by Becker & Newlin-Łukowicz (2018), have shown this to not be practically achievable to the degree asserted by folk belief.

1.3 England North-East accent features

Like New York, the North East of England is widely believed by locals to contain highly localised dialect variation. As Pearce (2009, 2011) shows, the Tyne and Wear (T&W) conurbation, centred on the cities of Newcastle and Sunderland, is divided perceptually at least into two dialects linked to Newcastle ("Geordie") and Sunderland ("Mackem"), although many locals will claim more fine-grained regional variation between even neighbouring towns is identifiable. However, there is relatively limited evidence of such within-T&W difference. For example, Burbano-Elizondo (2008a) finds that speakers from Sunderland participate in H-dropping to a greater extent than speakers from Newcastle, although both cities show lower rates than elsewhere. This feature, alongside the pronunciation of several vowels such as GOOSE, is commonly claimed by locals to distinguish Newcastle and Sunderland (Burbano-Elizondo 2008b, Pearce 2012). Pearce (2009) does suggest that such lay perceptions match patterns of local variation in production. However, with little research on linguistic production in the North East outside of Newcastle, it remains unclear whether differences are substantial, regional or class-based, and whether listeners can reliably perceive them (RQ2).

Given this, we chose four features claimed to vary within North East England: H-dropping and STAR at the county level; NURSE and GOAT at the borough level (Pearce 2009, Burbano-Elizondo 2008b). These may occur infrequently and mainly in production, but as perceptual variants they provide a starting point for testing how fine-grained local identification can be (RQ3). Following Pearce (2009) and Burbano-Elizondo (2008b), H-dropping is more common in southern counties (including Sunderland); STAR may occur as [ɑː] ∼ [ɒː] in Northumberland, [ɒː] in T&W, and [aː] in County Durham; NURSE may be [ɔː] in parts of Newcastle and [øː] in other T&W boroughs; GOAT may be [oː] in Sunderland vs. [ɵː] further north.

2 Methods

2.1 Gamified task design: stimuli, speakers, and procedures

2.1.1 Audio stimuli

The gamified task was designed to test folk belief about regional variation in North East England: that there is perceivable fine-grained geographic variation not only between Newcastle and Sunderland but town-to-town within the region, and that this geographical variation is a matter of kind (i.e., one town has a variant that another categorically does not) rather than rate (i.e., two towns share a pool of variants, but one town favours a particular variant in comparison to the other town). Audio stimuli were selected from two subcorpora of the Diachronic Electronic Corpus of Tyneside English (DECTE) (Corrigan et al. 2012): NECTE2 (the Newcastle Electronic Corpus of Tyneside English 2), recorded in 2007-2010; and another corpus recorded in 2023 (not-yet published). The audio corpora contained 203 speakers, who we grouped by county of origin (Northumberland, T&W, County Durham) and, within T&W, by borough (Newcastle, Gateshead, Sunderland, North Tyneside, South Tyneside). We note that Northumberland and County Durham are quite large, and include towns quite close to T&W (Ashington, Chester-le-Street, etc.) as well as towns quite removed from T&W (Berwick-upon-Tweed, Newton Aycliffe, etc.). However, if indeed the folk belief of fine-grained geographical difference bears out, a listener would be able to accurately classify speakers from Ashington or Berwick-upon-Tweed as being from Northumberland. Stimuli contained one of four sociolinguistic variables (Section 1.3): STAR vowel, H-dropping, GOAT vowel, and NURSE vowel. Note that our inclusion of these features is not us assuming a clear difference in production and testing whether listeners can recognize it. Rather, it should be taken as us giving participants a "fighting chance" of sorts at the task: if folk belief is that features such as these distinguish towns from one another by kind, and there is at least limited evidence that there may be geographical differences in rate for these features, then listeners would be most able to successfully identify speakers' regionality from stimuli such as these if the folk belief has merit.

Speakers were selected to balance speaker gender (female/male), age (below/above 30), and education level (below/above university level), resulting in 8 categories. Thus, stimuli at the county level came from 24 speakers (8 categories * 3 counties); at the borough level, from 40 speakers (8 categories * 5 boroughs). Each speaker appeared only once per level. Six speakers were used twice across the two levels as there was no available other speaker representing the specific category.

Two audio clips were extracted from each speaker, each (3-8s) containing a different target feature. For county-level, one utterance featured H-dropping (*stimuli_ h*) and the other the STAR vowel (*stimuli_ ar*). For borough-level, one utterance featured the NURSE vowel (*stimuli_ ur*) and the other the GOAT vowel (*stimuli_ ou*). Other features were not strictly controlled, to prioritise naturalness and authenticity.

2.1.2 Task levels

The gamified task asked players to identify the speaker's accent origin in two levels: a *county* level with 24 trials (Figure 1), and a *borough* level with 40 trials (Figure 2). In each trial, players manually played a one-time audio stimulus. Map options became clickable only after the audio had finished playing. Trials advanced automatically after each response, with no option to skip or return.

Fig. 1: Screenshot of the game interface at the county level.

To keep the gamified tasks concise, two versions were created, each combining one county-level and one borough-level feature. The *stimuli_ h_ ou* version combined county-level *stimuli_ h* and borough-level *stimuli_ ou*, while the *stimuli_ ar_ ur* version contained county-level *stimuli_ ar* and

Fig. 2: Screenshot of the game interface at the borough level.

borough-level *stimuli_ur*. There was no statistically significant difference in performance between the two versions ($\chi^2(1) = 3.31$, $p = 0.069$).

2.1.3 Gamified elements

We incorporated several gamified elements to maximise player engagement, while balancing these elements to safeguard the integrity of the results.

Multi-platform access: The game was built as an interactive map-clicking game using LabVanced (Finger et al. 2017). Players could access the game on mobile phones, tablets, or laptops, via a QR-code or a direct link. Before starting, they completed a short demographic questionnaire, designed to resemble the character-setting stage of a role-playing video game.

User interface and interaction: The game interface followed basic mobile-friendly UI/UX principles. It was visually engaging yet minimal (Figure 1 and Figure 2): a map, a compass, and a large play button, using a colour-blind friendly scheme. Maps were generated from UK Local Authority District digital boundary data (Office for National Statistics 2023) using ggplot2 (Wickham 2016) in R (R Core Team 2022).The map became clickable only after the audio finished. All clickable elements were mobile-appropriate, and location names were clearly displayed.

Progress bar: A progress bar tracked by-trial progress, as shown at the bottom of Figure 1 and Figure 2.

Points: Each trial was worth one point. Scores were shown twice: at the end of the county level and the borough level. Although displaying points after each trial could boost engagement, we avoided it to prevent perceptual learning (e.g., Lehet, Fenn & Nusbaum 2020).

Leaderboard: At the end of the game, a leaderboard appeared by the side of the player score, displaying the top three participants and their scores.

Rewards: The top three won shopping vouchers (£30, £20, £10), and four additional £10 vouchers were randomly awarded to other participants.

2.2 Game participants

Forty-seven out of 50 players (mean age = 31, SD = 9) who completed the game and self-identified as native English speakers were included in the analysis (Version *stimuli_h_ou*: N=21, Version *stimuli_ar_ur*: N = 26). Recruitment took place over two months via posters on campus and online mailing lists for phonetics and sociolinguistics modules (≈ 200 students). The data collection scale was small since this was a proof-of-concept test case. Figure 3 demonstrates participants' residence distribution: six participants neither currently reside in North-East nor lived the longest in North-East.

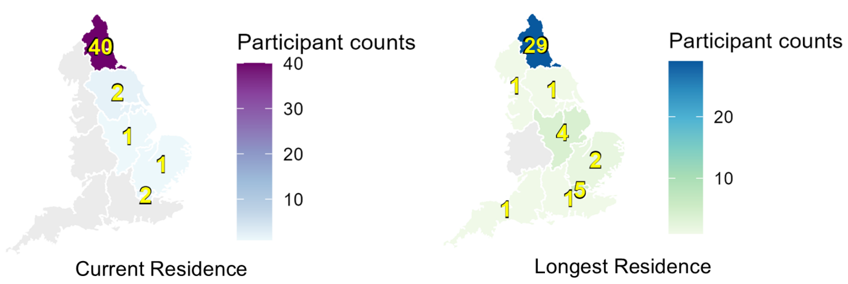

Fig. 3: Left: Participant (N=46) distribution by **current residence** in England. *Note: One participant self-reported living in France.* **Right**: Participant (N=44) distribution by **longest residence** in England. *Note: One participant did not report valid information. Two participants lived in Wales and Scotland.*

2.3 Data analyses

We analysed the data in two steps: exploratory and hypothesis-testing. First, random forest models (Breiman 2001) ranked predictors, trained separately at county and borough levels as an exploratory method to identify predictors of correct accent identification (CORRECT: 1 = correct, 0 = incorrect).

Predictors included listener-level (listener age, education, current, childhood and longest residence locations), speaker-level (origin, age, sex, and education), and stimulus-level variables (duration and word count). The dataset comprised 1,880 trials from 47 listeners, split into 70% training and 30% test sets using stratified sampling. Feature selection was performed using the `Boruta` package (Kursa & Rudnicki 2010); the final model was trained using only the confirmed important features. The `randomForest` package (Liaw & Wiener 2002) was used for model training, and evaluated using the `caret` package (Kuhn 2008).

Second, we conducted mixed-effects logistic regression models with `lme4` (Bates et al. 2015), focusing on Speaker Origin (the highest-ranking predictor) and our primary manipulation, Phonetic Feature. Fixed effects included both factors and their interaction; random intercepts were included for participant and item. Model selection used backward-comparison via likelihood ratio tests, with county- and borough-level models analysed separately. Pairwise comparisons used `emmeans` (Lenth 2024).

Additional exploratory analyses, based on the random forest results, are provided in our OSF repository.

3 Results and Discussion

3.1 Results

3.1.1 Descriptive statistics

At the county level, Figure 4 shows that participants were numerically more accurate on identifying speakers from T&W (49%) than those from Durham (37%) and Northumberland (30%). The confusion matrix (Figure 5) further indicates that participants showed a bias toward identifying speakers as being from T&W, with many Durham and Northumberland speakers incorrectly classified as T&W. This suggests that when uncertain, listeners defaulted to the most familiar or probable regional option.

At the borough level, speakers from Sunderland were identified with higher accuracy than other regions (Figure 6). Figure 7 further reveals that Tyneside boroughs (Newcastle, North Tyneside, South Tyneside, Gateshead) were often confused, with participants frequently misattributing speakers between these neighbouring regions. This pattern suggests that listeners recognised Sunderland's distinct features but struggled with Tyneside differences.

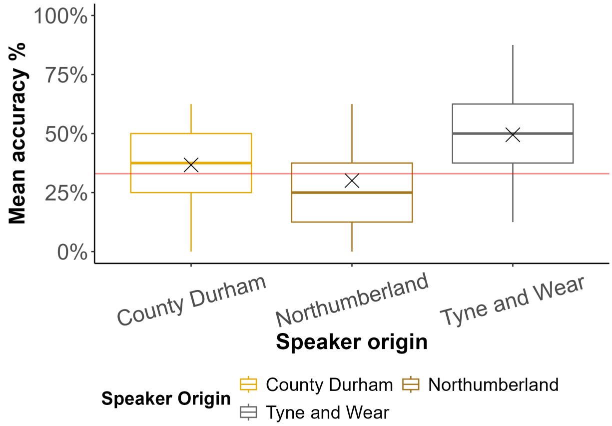

Fig. 4: Mean accuracy at the county level by SPEAKER ORIGIN (Durham, Northumberland, & T&W). Cross marks represent mean accuracy. The red y-axis intercept represents chance level at 33.3%.

3.1.2 Feature importance ranking

For the county level, four predictors were confirmed as important factors: SPEAKER_ORIGIN, STIMULI_DURATION, SPEAKER_EDUCATION, and WORD_COUNT. For the borough level, Boruta confirmed six important factors: SPEAKER_AGE, SPEAKER_EDUCATION, SPEAKER_ORIGIN, SPEAKER_SEX, STIMULI_DURATION and WORD_COUNT.

The final random forest models achieved lower overall accuracies than their no information rate (NIR) counterparts (i.e., when the majority class is always chosen) for both county and borough levels (county: 61.5%, NIR = 63.9%, p = .83; borough: 73%, NIR = 76.05%, $p < .005$). The below-chance performance suggests that the identification task was challenging and that listeners may have been using inconsistent cues or strategies. However, the variable importance rankings (Figure 8) remain informative for understanding which factors contributed most to the variation in responses, even in this difficult task. Where the speaker of the stimulus audio is from (SPEAKER_ORIGIN) is the most important variable for both levels. Stimuli-related factors such as length (duration and number of words) are important factors, which is not surprising, since statistical learning is positively correlated with the amount of data received. Speaker education is also an important feature for county-identification, together with the sex and age of the speaker in the borough level.

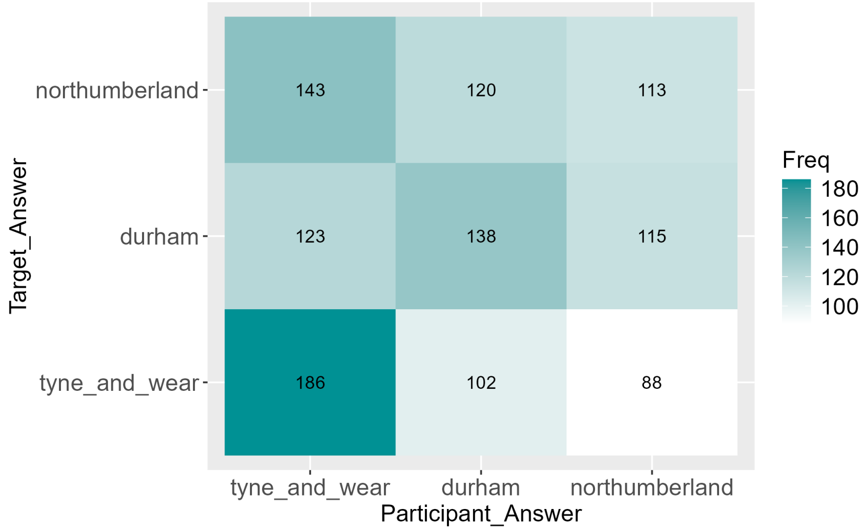

Fig. 5: Confusion matrix for county-level accent identification. Rows show speaker origins; columns show participant responses.

3.1.3 Effects of speaker origin and phonetic features

County-level model comparisons[1] confirmed that identification accuracy was higher on T&W than on Durham and Northumberland, $\chi^2(2) = 12.21$, $p = 0.002$. There was no effect of PHONETIC FEATURE, $\chi^2(1) = 0.92$, $p = 0.34$, or SPEAKER ORIGIN × PHONETIC FEATURE, $\chi^2(2) = 0.23$, $p = 0.89$.

Pairwise comparisons showed that recognition of T&W speakers was significantly more accurate than for Northumberland, $\beta = 0.88$, SE $= 0.25$, $p < .001$, and marginally more accurate than for Durham, $\beta = 0.55$, SE $= 0.24$, $p = .059$. Accuracy did not differ significantly between speakers from Northumberland and Durham, $\beta = -0.34$, SE $= 0.25$, $p = .36$.

Borough-level model comparisons[2] confirmed that speakers from Sunderland were identified with higher accuracy than other regions (Figure 6), $\chi^2(4) = 12.52$, $p = 0.01$. There was no effect of PHONETIC FEATURE, $\chi^2(1) = 1.47$, $p = 0.23$, or PHONETIC FEATURE x SPEAKER ORIGIN, $\chi^2(4) = 2.48$, $p = 0.65$.

1 Best-fit model: *glmer(correct ∼ accent_origin + (1|item) + (1|participant))*.
2 Best-fit model: *glmer(correct ∼ accent_origin + (1|item) + (1|participant))*

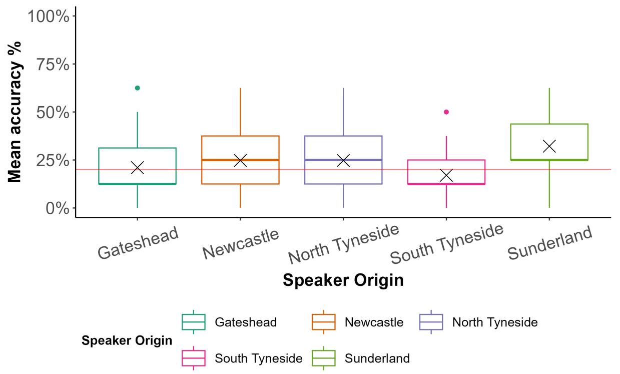

Fig. 6: Mean accuracy at the borough level by SPEAKER ORIGIN (Gateshead, Newcastle, North Tyneside, South Tyneside, & Sunderland). Cross marks represent mean accuracy. The red y-axis intercept represents chance level at 20%.

Pairwise comparisons indicated that the only significant difference in identification accuracy was between speakers from Sunderland (32%) and South Tyneside (17%), $\beta = 0.85$, $SE = 0.24$, $p = 0.004$.

3.2 Discussion on accent identification

This study examined county- and borough-level accent identification in North East England, testing the folk belief that such distinctions are possible. We investigated whether recognition depended on speaker origin or phonetic features (i.e., GOAT, H-dropping, STAR, and NURSE). Some signal emerged: T&W speakers were more identifiable, likely due to local accents familiarity (Kerswill & Williams 2002, Leach, Watson & Gnevsheva 2016), as more than half of the participants were from T&W. At the borough level, Sunderland speakers were most accurately recognised, despite most listeners being from Newcastle, a pattern consistent with the recognisability of culturally prominent areas (Leach, Watson & Gnevsheva 2016). Additionally, the four phonetic features did not predict recognition effectively, perhaps because short audio clips (around 10s) were stripped of context and limited access to broader linguistic cues. This indicates that it can be difficult for listeners to extract social associations from other linguistic aspects (lexicon, syntax, pragmatics, etc.) apart from the accent itself. However, note that Becker & Newlin-Łukowicz (2018) found that although listeners could not accurately place speakers within New York City, they nevertheless placed

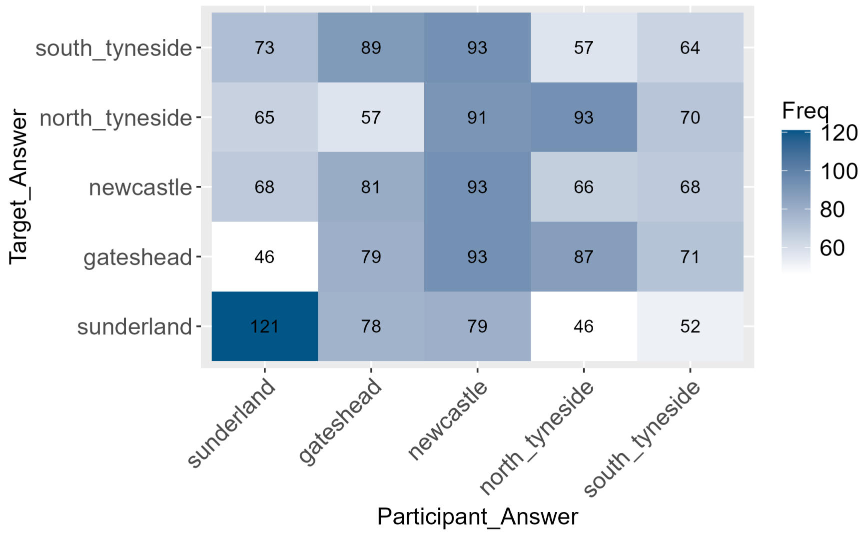

Fig. 7: Confusion matrix for borough-level accent identification. Rows show speaker origins; columns show participant responses.

speakers strategically based on other perceived social categories and attributes. Similarly, our results may reflect listeners' knowledge of how class and geography intersect in North East England. Indeed, speaker education level being a key factor influencing accuracy in the county-identification level suggests that listeners may be mapping demographic, rather than linguistic, features onto their mental map of the region. The limited signal may therefore reflect not an accurate folk belief in fine-grained geographical differences in kind of linguistic production, but instead listeners' mappings of social indexicalities within the region.

4 General Discussion

We demonstrated that a gamified design can successfully replace traditional per-participant incentives as a data collection strategy in a phonetic perception study. For researchers, data collection was efficient in terms of both recruitment and cost. Most data in this study were obtained by advertising the game through class mailing lists and via physical posters on campus. Although the final number was not as big as those large-scale studies, com-

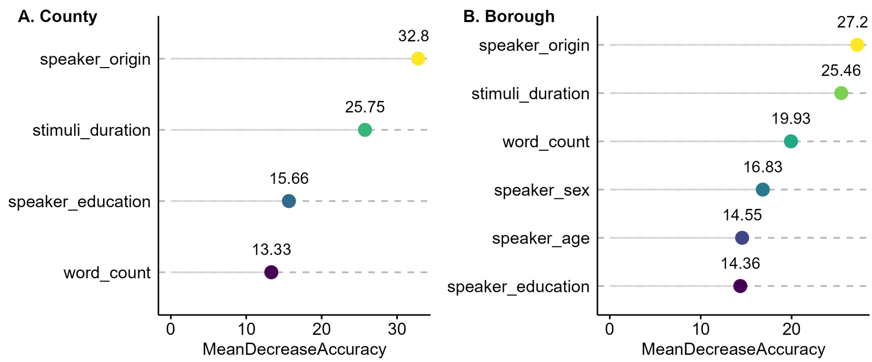

Fig. 8: Variable importance plots based on mean decrease in accuracy (higher = more important). A: top county-level predictors; B: top borough-level predictors.

pared with traditional laboratory experiments, recruitment required little active encouragement, as many participants were willing to take part simply because the task was presented as a game. We spent £100 on vouchers in total. By contrast, a traditional experiment would require £300–£600 in participant payments at the current minimum wage (£12.21/hour; UK Government 2024), while likely yielding comparable or smaller datasets. Participants also reported that the game felt less "intimidating" than a traditional lab task. The sociophonetic test case was especially well-suited to gamification, as it does not require precise playback timing measures or accurate reaction time records. With the topic's natural appeal to the public, gamifying the experiment can further support engagement with citizen science through knowledge-exchange events. Using an improved version of the game, we subsequently collected additional data during a 5-hour ESRC Festival of Social Science event and a 6-hour Discover Festival pop-up stand, attracting over 100 and around 50 participants, respectively.

However, implementing gamified studies, especially with customised digital games, is not without hurdles. For a gamified design with an attractive user interface and interaction, a digital platform is needed. For traditional experiments, free tools such as PsyToolkit (Stoet 2017), PsychoPy (Peirce 2007) and jsPsych (De Leeuw 2015) are available, and there are plenty of readily built sample tasks. However, these platforms are less flexible when it comes to the design mechanisms and user interactivity demanded by gamified approaches. Moreover, there are few freely available options and they are not straightforward to incorporate without the necessary skill sets — such as programming and UX design — considering that gamification is usually applied in other sectors (e.g., social sciences) than computer science.

The same issue occurs when hosting a task on free servers (e.g., university servers). Especially with the function of data collection, coding a gamified task from scratch requires strong programming skills, and good infrastructure support. Another way of developing tasks is using hosting platforms that include game building abilities such as LabVanced, Gorilla, and Genially[3]. However, these platforms often run on a subscription basis and once the subscription runs out, the game is no longer accessible. Nevertheless, it is encouraging to see that some of these experiment-building platforms have started to incorporate game-builders (e.g., Gorilla).

Gamification needs to go beyond explicit employment of individual gamified elements (e.g., the points-badges-leaderboard design (PBL) (Escher 2016, Seaborn & Fels 2015). These elements can be effective if the tasks are progress-oriented, such as learning or behaviour tracking (e.g Shortt et al. 2023, Sanchez, Langer & Kaur 2020, Smith, Legaki & Hamari 2022), or if the rules are simple, which helps to keep games short and maintain attention. To employ gamified elements in such contexts is straightforward, and the effect can potentially be measured in a controlled experiment. However, issues arise when game rules get complicated or when large-scale repeated-measures are required (Schöbel et al. 2020). For instance, in typical psychological experiments involving repeated trials over a long time, simple implementation of the gamified elements would not work to reduce fatigue effect, as participants get bored naturally when the game length reaches a certain point (Seaborn & Fels 2015, Landers & Landers 2014). Although the current project managed to collect simple click responses through gamification, a majority of phonetic research require high-quality production data or more complex behavioural responses which may involve complex game mechanics. In this sense, gamified elements beyond explicit structural components might be more helpful. Without explicit employment of the PBL design, the logic-based word ladder game (Genovese et al. 2024) or the narrative-based the Iowa Gambling Tasks (IGT) (Brevers et al. 2013) are examples of competition- and engagement-oriented gamified forms.

3 https://www.labvanced.com/, https://gorilla.sc/, https://genially.com/

5 Conclusion

This study demonstrates how gamification can be incorporated into phonetic research, offering both methodological innovation and practical insight. By detailing the benefits and challenges encountered, we aim to provide a useful reference point for researchers considering similar approaches. Our sociophonetic findings challenge the folk assumption that regional dialects are easily recognisable, revealing that local residents often struggle to identify speakers' origins. We hope this work encourages further exploration of playful, participatory citizen science methods in phonetics and sparks new discussions around dialect perception.

6 Data Availability Statements

The data are available at the Open Science Framework (OSF) at https://doi.org/10.17605/OSF.IO/A7S3B.

7 Acknowledgements

This project was funded by the Institute for Social Science at Newcastle University through a Pioneer Award (OSR/0352/NISR/0002) awarded to Cong Zhang. The subsequent knowledge exchange events, *Wey aye, man: Think you know a North-East accent when you hear one?* (Festival of Social Science) and *Howay! Think you know a North-East accent when you hear one?* (Discover Festival, Newcastle University and the Great North Museum: Hancock), were supported by the Economic and Social Research Council (ESRC) through an Impact Acceleration Account (IAA) and a Discover Festival grant, respectively.

References

Arvaniti, Amalia, Stella Gryllia, Cong Zhang & Katherine Marcoux. 2022. Disentangling emphasis from pragmatic contrastivity in the English H*˜ L+ H* contrast. In *Proceedings of Speech Prosody 2022*, 837–841. https://doi.org/10.21437/SpeechProsody.2022-170.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Becker, Kara & Luiza Newlin-Łukowicz. 2018. The myth of the New York City borough accent: Evidence from perception. *University of Pennsylvania Working Papers in Linguistics* 24(2). Article 3. https://repository.upenn.edu/pwpl/vol24/iss2/3.

Bertolo, Mila, Martynas Snarskis, Thanos Kyritsis, Lidya Yurdum, Constance M. Bainbridge, S. Atwood, Courtney B. Hilton, Anya Keomurjian, Judy S. Lee, Alex Mackiel, Vanessa Mak, Mijoo Shin, Alma Bitran, Dor Shilton, Lana Delasanta, Hang Do, Jenna Lang, Tenaaz Irani, Jayanthini Kangatharan, Kevin Lafleur, Nashua Malko, Quentin D. Atkinson, Manvir Singh & Samuel A. Mehr. 2025. The expanded natural history of song discography, a global corpus of vocal music. *Open Mind* 9. 844–863.

Bozkurt, Aras & Gürhan Durak. 2018. A systematic review of gamification research: In pursuit of homo ludens. *International Journal of Game-Based Learning (IJGBL)* 8. 15–33. https://doi.org/10.4018/IJGBL.2018070102.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32.

Brevers, Damien, Antoine Bechara, Axel Cleeremans & Xavier Noël. 2013. Iowa Gambling Task (IGT): Twenty years after–gambling disorder and IGT. *Frontiers in Psychology* 4. 665.

Burbano-Elizondo, Lourdes. 2008a. *Language variation and identity in Sunderland*. Sheffield: University of Sheffield dissertation.

Burbano-Elizondo, Lourdes. 2008b. Regional variation and identity in Sunderland. In Tope Omoniyi & Goodith White (eds.), *The Sociolinguistics of Identity*, 113–128. New York: Bloomsbury.

Cho, Eun, Lidya Yurdum, Ekanem Ebinne, Courtney B. Hilton, Estelle Lai, Mila Bertolo, Pip Brown, Brooke Milosh, Haran Sened, Diana I. Tamir & Samuel A. Mehr. 2025. Ecological momentary assessment reveals causal effects of music enrichment on infant mood. *Child Development*. 1555–1567. https://doi.org/https://doi.org/10.1111/cdev.14246.

Corrigan, Karen P., Isabelle Buchstaller, Adam Mearns & Hermann Moisl. 2012. The Diachronic Electronic Corpus of Tyneside English. Newcastle University. https://research.ncl.ac.uk/decte.

Coupland, Nikolas & Hywel Bishop. 2007. Ideologised values for British accents. *Journal of Sociolinguistics* 11(1). 74–93. https://doi.org/10.1111/j.1467-9841.2007.00311.x.

De Leeuw, Joshua R. 2015. Jspsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods* 47. 1–12.

Deterding, Sebastian, Dan Dixon, Rilla Khaled & Lennart Nacke. 2011. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: envisioning Future Media Environments*, 9–15.

Duncan, Daniel. 2019. The influence of suburban development and metropolitan fragmentation on language variation and change: Evidence from Greater St. Louis. *Journal of Linguistic Geography* 7(2). 82–97. https://doi.org/10.1017/jlg.2019.8.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12. 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x.

Eryiğit, Gülşen, Ali Şentaş & Johanna Monti. 2023. Gamified crowdsourcing for idiom corpora construction. *Natural Language Engineering* 29(4). 909–941.

Escher, MC. 2016. Towards meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly*. https://doi.org/10.25300/MISQ/2017/41.4.01.

Finger, Holger, Caspar Goeke, Dorena Diekamp, Kai Standvoß & Peter König. 2017. Labvanced: A unified JavaScript framework for online studies. In *International Conference on Computational Social Science (Cologne)*, 1–3.

Genovese, Francesca, Marianna Marcella Bolognesi, Angelo Di Iorio & Fabio Vitali. 2024. The advantages of gamification for collecting linguistic data: A case study using Word Ladders. *Online Journal of Communication and Media Technologies* 14(2). 1–17.

Hamari, Juho, Jonna Koivisto & Harri Sarsa. 2014. Does gamification work?–A literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences*, 3025–3034.

Hartshorne, Joshua K., Joshua B. Tenenbaum & Steven Pinker. 2018. A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition* 177. 263–277. https://doi.org/10.1016/j.cognition.2018.04.007.

Hilton, Nanna H. 2021. Stimmen: a citizen science approach to minority language sociolinguistics. *Linguistics Vanguard* 7(s1). 20190017.

Kerswill, Paul & Ann Williams. 2002. Dialect recognition and speech community focusing in new and old towns in England: The effects of dialect levelling, demography and social networks. In Daniel Long & Dennis Preston (eds.), *Handbook of Perceptual Dialectology*, vol. 2. Amsterdam: John Benjamins. https://doi.org/10.1075/z.hpd2.16ker.

Kim, Bohyun. 2015. Designing gamification in the right way. *Library Technology Reports* 51(2). 29–35.

Kim, Yoolim, Vita V. Kogan & Cong Zhang. 2023. Collecting big data through citizen science: Gamification and game-based approaches to data collection in applied linguistics. *Applied Linguistics*. amad039. https://doi.org/10.1093/applin/amad039.

Kogan, Vita V. 2025. Ahead of the game: applying gamification and game-base teaching to linguistics. In *Teaching linguistics with games*, 1–11. Routledge.

Kuhn, Max. 2008. Building predictive models in R using the *caret* package. *Journal of Statistical Software* 28. 1–26.

Kursa, Miron B. & Witold R. Rudnicki. 2010. Feature selection with the Boruta package. *Journal of Statistical Software* 36. 1–13.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Landers, Richard N., Elena M. Auer, Andrew B. Collmus & Michael B. Armstrong. 2018. Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming* 49. 315–337. https://doi.org/10.1177/1046878118774385.

Landers, Richard N. & Amy K. Landers. 2014. An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming* 45(6). 769–785.

Leach, Hannah, Kevin Watson & Ksenia Gnevsheva. 2016. Perceptual dialectology in northern England: Accent recognition, geographical proximity and cultural prominence. *Journal of Sociolinguistics* 20(2). 192–211. https://doi.org/10.1111/josl.12178.

Leemann, Adrian, Curdin Derungs & Stephan Elspaß. 2019. Analyzing linguistic variation and change using gamification web apps: The case of German-speaking Europe. *PLOS ONE* 14(12). e0225399. https://doi.org/10.1371/journal.pone.0225399.

Leemann, Adrian, Marie-José Kolly & David Britain. 2018. The english dialects app: the creation of a crowdsourced dialect corpus. *Ampersand* 5. 1–17. https://doi.org/10.1016/j.amper.2017.11.001.

Leemann, Adrian, Marie-José Kolly, Ross Purves, David Britain & Elvira Glaser. 2016. Crowdsourcing language change with smartphone applications. *PLoS ONE* 11(1). e0143060. https://doi.org/10.1371/journal.pone.0143060.

Lehet, Matthew I., Kimberly M. Fenn & Howard C. Nusbaum. 2020. Shaping perceptual learning of synthetic speech through feedback. *Psychonomic Bulletin & Review* 27(5). 1043–1051.

Lenth, Russell V. 2024. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.3. https://CRAN.R-project.org/package=emmeans.

Liaw, Andy & Matthew Wiener. 2002. Classification and regression by randomForest. *R News* 2(3). 18–22.

Long, Bria, Jan Simson, Andrés Buxó-Lugo, Duane G. Watson & Samuel A. Mehr. 2023. How games can make behavioural science better. *Nature* 613(7944). 433–436.

Office for National Statistics. 2023. *UK Local Authority Districts (May 2023) UK BFC*. Accessed: 2023-05-30. Office for National Statistics. https://geoportal.statistics.gov.uk/datasets/ons::local-authority-districts-may-2023-boundaries-uk-bfc/.

Pearce, Michael. 2009. A perceptual dialect map of North East England. *Journal of English Linguistics* 37(2). 162–192. https://doi.org/10.1177/0075424209334026.

Pearce, Michael. 2011. Exploring a perceptual dialect boundary in North East England. *Dialectologia et Geolinguistica* 19(1). 3–22. https://doi.org/10.1515/DIG.2011.001.

Pearce, Michael. 2012. Folk accounts of dialect differences in Tyne and Wear. *Dialectologia et Geolinguistica* 20(1). 5–25. https://doi.org/10.1515/DIG.2012.001.

Peirce, Jonathan W. 2007. Psychopy—psychophysics software in Python. *Journal of Neuroscience Methods* 162(1-2). 8–13.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ryan, Richard M. & Edward L. Deci. 2000a. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* 25(1). 54–67.

Ryan, Richard M. & Edward L. Deci. 2000b. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55(1). 68.

Sanchez, Diana R., Markus Langer & Rupinder Kaur. 2020. Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Computers & Education* 144. 103666. https://doi.org/10.1016/j.compedu.2019.103666.

Schöbel, Sofia, Andreas Janson, Katharina Jahn, Bastian Kordyaka, Ozgur Turetken, Naza Djafarova, Mohammed Saqr, Dezhi Wu, Matthias Söllner, Martin Adam, Povl H. Gad, Henrik Wesseloh & Jan M. Leimester. 2020. A research agenda for the why, what, and how of gamification designs: Outcomes of an ECIS 2019 panel. *Communications of the Association for Information Systems* 46. 706–721.

Seaborn, Katie & Deborah I. Fels. 2015. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies* 74. 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006.

Shahzad, Muhammad F., Shuo Xu, Obaid ul Rehman & Iqra Javed. 2023. Impact of gamification on green consumption behavior integrating technological awareness, motivation, enjoyment and virtual CSR. *Scientific Reports* 13(1). 21751.

Shortt, Mitchell, Shantanu Tilak, Irina Kuznetcova, Bethany Martens & Babatunde Akinkuolie. 2023. Gamification in mobile-assisted language learning: a systematic review of duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning* 36(3). 517–554.

Smith, Annique, Zampeta Legaki & Juho Hamari. 2022. Games and gamification in flipped classrooms: A systematic review. *International GamiFIN Conference*. 33–43.

Stoet, Gijsbert. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44(1). 24–31.

Thompson, Henry S., Anne H. Anderson, Ellen G. Bard, Gwyneth Doherty-Sneddon, Alison Newlands & Cathy Sotillo. 1993. The HCRC map task corpus: Natural dialogue for speech recognition. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993*.

UK Government. 2024. *National Minimum Wage and National Living Wage Rates*. Accessed: 2025-07-21. https://www.gov.uk/national-minimum-wage-rates.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Zichermann, Gabe & Christopher Cunningham. 2011. *Gamification by design: Implementing game mechanics in web and mobile apps*. Sebastopol, CA: O'Reilly Media.