JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Speaker evaluation across the adult lifespan: Examining age and gender effects in the perception of Tyneside voices

Johanna Mechler[1]*
[1]University of Duisburg-Essen, Germany

Abstract. Research on the sociolinguistic monitor (Labov et al. 2011) suggests that evaluative reactions to voices are influenced by social and linguistic characteristics of the speakers and listeners. To date, however, we know very little about the impact of different stages of adulthood on linguistic perception. This study explores how the perception of vernacular speech is mediated by speaker and listener traits, particularly by age and gender. Sixteen speech samples from sociolinguistic interviews (Buchstaller 2024) conducted in the Tyneside region of the UK were manipulated acoustically for variants of (ing). Listeners (n = 72) rated speakers for status and social likability attributes and were asked to estimate speakers' ages. Regression analyses revealed three key findings: First, the realization of (ing) across guises was only minimally important to listener evaluations of speakers in the context of stimuli-rich guises. Second, evaluations are highly contingent on speaker age and gender — listeners rate male vernacular speakers in their 40s higher on professionalism and intelligence, while younger female speakers are perceived as significantly more friendly and trustworthy. Finally, older male respondents are more critical in their evaluations, i.e., they assign lower ratings, than younger female listeners (Levon et al. 2021). Overall, this study highlights the interactional nature of age and gender in sociolinguistic perception, offering a nuanced understanding of how age- and gender-related stereotypes affect speaker evaluation.

Keywords. sociolinguistic perception, lifespan, age, gender, vernacularity, Tyneside

A peer-reviewed contribution to the *Journal of Language and Aging Research* (JLAR).
Submitted: 2024-08-13
Accepted: 2025-05-23
Published: 2025-06-09
DOI: 10.15460/jlar.2025.3.1.XXX

*. Corresponding author, johanna.mechler@uni-due.de

Please cite as Mechler, Johanna. 2025. "Speaker evaluation across the adult lifespan: Examining age and gender effects in the perception of Tyneside voices." *Journal of Language and Aging Research* 3(1): 48–75. 10.15460/JLAR. 2025.3.1.XXX.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

1 Introduction

A fundamental concept in sociolinguistics is age-grading, a pattern of retrenchment whereby, generation after generation, individuals modify their linguistic behavior throughout their lifetime, resulting in a "regular association of a variant with a particular life-stage" (Wagner 2012a, 371). Age-graded variability has been hypothesized typically for stable variables as a response to linguistic marketplace pressures (Bourdieu and Boltanski 1975). Throughout the lifespan it is assumed to result in a "u-shaped" pattern consisting of a peak in vernacularity in adolescence, a middle-aged trough representing a decrease in vernacular variants (Wagner 2012b), followed by an orientation back towards higher vernacular rates in older age (Downes 1998). The decline in vernacular forms is undergirded by socio-psychological research: Arnett's (2000, 2007) research supports the claim that speakers feel and react to socio-normative pressures to use more standard variants to present themselves as professionals, when they enter the workforce in early to middle adulthood (see 2.2).

The expectation for speakers to become linguistically more "standard" as they move from adolescence into adulthood has been explored in an increasing number of studies, which have also documented the complexity of this transition. Indeed, panel studies have shown that speakers may take multiple trajectories as they reach "full engagement in the work force and family responsibilities" (Labov 2001, 101). Some individuals, particularly those in language sensitive occupations, follow the "u-shaped curve", i.e., after a vernacular spike in adolescence, they show retrenchment during their professional years and then revert to vernacularity once they reach retirement (e. g., Grama et al. 2023; Wagner 2012b). These speakers seem to be sensitive to socio-normative pressures based on hegemonic expectations. However, panel research has also demonstrated that this pattern is far from universal. Other individuals diverge from this pattern and maintain their vernacular rates, even as they enter professional careers (e. g., Moelders 2025; Beaman 2021). Hence, panel research has started to build an increasingly complex picture of the contingent ways in which speakers behave linguistically across their lifespans. What we lack, to date, is perception data on how these speakers and the linguistic features they produce in spontaneous speech are evaluated at different (adult) life stages. Given age-graded expectations, are listeners more permissive when hearing younger speakers produce vernacular forms? Conversely, are middle-aged speakers rated more negatively when using vernacular features? How do these perceptions interact with speaker and listener gender? To find answers to these questions, a matched-guise experiment was conducted using alveolar and velar (ing) variants in audio samples from sociolinguistics interviews, capturing "naturalistic" speech from Tyneside speakers in their 20s or 40s. This approach allows me to add fine-grained perceptual detail to the analysis of age-specific linguistic patterns.

Methodologically, previous perception studies have relied on highly controlled guises of (ing) (e. g., Levon and Fox 2014; Labov 2001). The guises used in this study are derived from naturally occurring, vernacular speech in sociolinguistic interviews. This study design thus offers insight into more complex situated perceptions than the controlled environments, which are usually used, do. Hence, I will consider both holistic and more focused reactions to variants of (ing). Further, one of the contributions this paper aims to make is to consider the effects of the listener on evaluative responses. Broadly, the listeners in this study come from English-speaking regions,

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

the US and the UK. There is a general consensus among perception work that older listeners are more critical towards vernacular speakers than younger listeners (Mechler, in press; Levon et al. 2021; Labov et al. 2011; Ball 1983; Giles 1970). What we lack is full understanding of whether these patterns emerge due to internalized marketplace pressures that increase with age; we also do not know whether informant age interacts with other listener traits such as gender or regional attitudes. It is this very aspect—the impact of socio-demographic factors on prescriptive pressures—where this study intends to establish a link between the linguistic malleability across the lifespan we see in production and the observed age effects at specific life stages in perception. In summary, the current study seeks to answer the following three research questions:

1. Do vernacular voices affect perceived professionalism / intelligence / friendliness / trustworthiness? (RQ1)
2. How are listener perceptions mediated by speaker characteristics of age and gender? In other words, what is the social costliness of speaker vernacularity at different life stages? (RQ2)
3. How are listener perceptions mediated by listener characteristics of region, age, and gender? (RQ3)

By highlighting the socially differentiated nature of perception of naturally occurring speech, this study aims to shed light on the intricate details of the effects of speaker age and gender as well as listener age and gender.

2 Background

Perception research not only demonstrates that listeners can evaluate speaker age and gender accurately, but that older and younger voices are evaluated differently and that the acoustic correlates of chronological and/or perceived age and gender carry social meaning(s) (e. g., Stewart and Ryan 1982; Montepare, Kempler, and McLaughlin-Volpe 2014; Lavan and McGettigan 2023). For (ing), previous studies (e. g. Campbell-Kibler 2008; Labov et al. 2011) show that manipulating the variable can influence listener evaluations and consequently provide insight into age-grading. However, another possibility is that the impact of (ing) is minimal compared to perceived speaker age and gender, especially in stimuli-rich contexts. In this study, I will test the importance of phonological variables against the effect of social / speaker qualities in spontaneous speech, highlighting possible tensions between expected and actual salience of (ing).

2.1 Operationalizing age and life stages

Throughout our life history we move through different ontogenetic stages, which are each characterized by globally different roles and expectations (Arnett 2000, 2007); these transitions have long been theorized to have an effect on linguistic choices (Downes 1998 and following). Eckert (1997) argued that speakers are exposed to language norms, increasingly so as they enter adulthood. At this life stage, speakers become part of the workforce and the linguistic marketplace, where normative

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

pressures are acute (Bourdieu and Boltanski 1975). There is some evidence (Arnett 2000, 2007; Bourdieu and Boltanski 1975) that speakers' awareness of societal and linguistic expectations can result in more standard speech as individuals move into adulthood (Grama et al., in press; Guimarães 2014; Labov 2001).

Due to great individual variability, "chronological age does not contain [absolute] information on an individual's development over time" (Schmitt and Unger 2019, 372). Indeed, Pfenninger (2024) notes that when exactly speakers reach a certain life stage might vary. On the basis of these findings, this study operationalizes a rather loose concept of life stages. I aim to test the perception of speakers at two life stages: early adulthood (Arnett 2000), represented by four 20-year-old speakers, and middle adulthood with speakers' "full engagement in the work force /family responsibilities" (Labov 2001, 101), represented by four 40-year-olds. This later stage follows after the career rising years (Buchstaller and Traugott 2006; Guimarães 2014), which speakers typically experience in their 30s. However, it is not only speaker age that might impact perception, but also listener age, which I will explain next.

2.2 Effects of listener traits on perception

Perception studies have accumulated a body of evidence on the interaction of age and social evaluations. These studies converge on the finding that older listeners seem to be more critical than younger respondents. Starting already in adolescence, Giles (1970, 219) finds that a "17-year-old is more "old-fashioned" [...] than the 12-year-old", meaning that the older teenagers rate speakers more conservatively and more critically. Similarly, Ball (1983, 170) reports that "late adolescents to some extent accorded more prestige to Received Pronunciation than early adolescents". Labov et al. (2011) also find that younger participants (ages 18-23) either only show a linear reaction or no reaction at all to different frequencies of (ing) variants compared to the logarithmic responses of older participants (ages 23-31). The few studies conducted on middle-aged and older listeners further confirm this pattern. Levon et al. (2021, 369) describe an age split at 45 years, with older respondents reacting more critically than younger ones, irrespective of their social status. Mechler (in press) perception study corroborates this finding, reporting that older listeners (ages 41-75) assign lower professionalism ratings than the younger listener groups (ages 18-26 and 27-40). Based on these studies, we can postulate two trends about listener age and sociolinguistic perception: first, listeners of different ages vary in how they evaluate speaker voices, and second, when comparing two age groups, the older respondents tend to be more judgmental than their younger counterparts. We do not fully understand yet if these observed patterns emerge due to internalized pressures that increase with age (see 2.1; Arnett 2000, 2007; Bourdieu and Boltanski 1975). The following analysis of the relationship between listener age and evaluative ratings tests the trends reported in previous perception research on the basis of a sample that cross correlates speaker and hearer age.

2.3 Linguistic variable (ing)

The linguistic variable that will serve as the current test case is the sociolinguistic marker (ing) — a well-established stable linguistic variable. In all speech communities where it has been studied, (ing) has at least two forms: the "standard" velar (e. g., *I am baki*[ŋ]) and the "vernacular" alveolar variant (e. g., *I am baki*[n]). Apparent time

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

research has identified (ing) as diachronically stable[2] (Wagner 2012b, 183) in varieties of English worldwide, being consistently conditioned by a wealth of social and linguistic factors (Tamminga 2019; Schleef, Meyerhoff, and Clark 2011; Hazen 2006; Labov 2001; Trudgill 1974). The age-graded nature of (ing) has most recently been investigated in panel studies. Wagner (2012b) demonstrates that students anticipate the move towards prestigious Ivy League universities by producing standard velar variants at elevated rates, in comparison to those moving on to state schools. Mechler and Buchstaller (2019) investigate middle-aged and older speakers, suggesting that middle class-oriented "social risers" show evidence of retrenchment to standard forms, while stably working class-oriented speakers are less prone to moving away from the vernacular. This pattern is supported by Grama et al. (2023), whose analysis of (ing) reveals that the u-shaped curve is occupationally niched; only speakers impacted by the *marché scolaire*, i.e., those working as professional educators, demonstrate clear retrenchment during their working years, followed by a tail in retirement. In terms of perception, the "standard" velar variant is unsurprisingly associated with higher prestige and the vernacular alveolar variant with lower prestige. Notably these evaluations have been shown to be contingent on the region and social class of the listener (e. g., Schleef and Flynn 2015; Campbell-Kibler 2008): Studies conducted in the US by Labov et al. (2011) and Campbell-Kibler (2011a, 2011b, 2009) confirm that speakers using the [ŋ] variant are perceived as more reliable, posher and more articulate, thus as more professional. Conversely, the social meaning of alveolar [n] is linked to Southern accents and a "lack of education, the country, and the term 'redneck'" (Campbell-Kibler 2007, 32). Research with UK listeners has long reported similar patterns in that the velar variant is perceived as the "standard", more prestigious variant. But more recent studies uncover more differentiated associations, suggesting that the evaluative difference between the two variants might be less pronounced and indeed regionally varied (Schleef, Flynn, and Barras 2017). Especially in the North of the UK, the difference in the evaluation of [n] and [ŋ] appears to be much more attenuated. More specifically, the alveolar variant seems to be less salient and less stigmatized than in the US (Levon and Fox 2014; Mechler, in press). My study adds nuance to this discussion, while supporting these findings.

2.4 Perception of Tyneside English

The variety in focus is Tyneside English (or, as it is locally known, Geordie), which is spoken in the North-East of England (see Beal, Burbano-Elizondo, and Llamas 2012; Montgomery 2007; Wells 1982). The Tyneside conurbation has been characterized by the rise and fall of heavy industry, which has given way to emerging economies in higher education and sustainable energy, especially in larger cities like Newcastle and Gateshead (Beal, Burbano-Elizondo, and Llamas 2012; Vall 2007). Due to the historical working-class background of many Tyneside families, Geordies tend to take great pride in their heritage and culture, including their linguistic heritage (Beal, Burbano-Elizondo, and Llamas 2012). The patterns of linguistic production in the North-East are well described as the variety continues to be an object of study (Grama et al. 2023; Mechler et al. 2022; Moelders 2025; Buchstaller et al. 2017; Fehringer and Corrigan 2015; Childs 2013; Watt 2002; Wells 1982). Typical for Tyneside English

2. Note, however, that the few available trend studies have shown subtle shifts in its linguistic and /or social conditioning (e. g., Forrest and Wolfram 2019; Travis, Grama, and Purser 2023).

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

are the following features: /t/- and /k/-glottalization, FACE and GOAT monophthongization, T-to-R, localized forms of negation (*divven't* / *dinnet* for *don't* / *didn't*), and the absence of a FOOT∼STRUT split (Watt 2000; Watt and Milroy 1999 inter alia). The ample research on language production and, increasingly, on lifespan change in the Tyneside area contrasts with a scarce—albeit growing—number of perception and attitude studies. From an outsider's or non-Geordie perspective, perception research documents strong language attitudes that British people harbor towards Northern English varieties (Coupland and Bishop 2007; even if less accentuated in hiring contexts, see Levon et al. 2021). For example, McKenzie and McNeill (2022) find that while listeners express more favorable language attitudes when asked explicitly, implicitly they still hold strong negative prejudices against Northern English speakers. In their study on listeners' sensitivity to phonetic versus grammatical variability, Levon and Buchstaller (2015, 336) show that "increased frequency of [the typically Northern feature] NSR [(Northern Subject Rule)] causes listeners to judge a speaker as sounding less professional,' particularly in the overtly prestigious context of a newscast". This finding is corroborated by Levon, Buchstaller, and Mearns (2020) work which reveals that speakers with typically Northern variants are downgraded, whereas typically Southern (and thus more standardized) variants are not penalized in this way. Notwithstanding these findings, Mechler (in press) describes an interesting age effect in such perceptions: Re-recordings of the same Tyneside speakers at different ages reveals that 30-year-old speakers are perceived as more professional regardless of their degree of vernacularity.

3 Methods

The present study relied on a matched-guise experiment (Lambert et al. 1960). Participants listened to sound samples of the same speakers with different guises (with [n] or [ŋ]) and evaluated them for status (professionalism, intelligence) and social likability (friendliness, trustworthiness) traits. All listeners heard all samples as part of a within-subject study design. The following sections illustrate the methodology in more detail.

3.1 Data collection and study design

Listeners were recruited from two broad geographical locations, namely the US and the UK. All data collection in the US took place in person in the "Language Pod", which is a language lab, where different kinds of linguistic research are conducted as part of a cooperation between The Ohio State University and a local science museum. The UK data comes from in-person data collection in a language lab at Newcastle University and online data collection.[3] In data analysis, the factor of type of data collection (in person / online) was not selected as significant in the model building process, so it is assumed that for this experiment the setting did not have an effect on the results. In both locales, the study consisted of three parts: first, a block of basic sociodemographic questions, then the matched-guise experiment itself, followed by

3. As the experiment required a computer and headphones, participants likely completed the experiment at home; however, the exact location / setting was not part of the questionnaire, so it might have varied.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

regional and attitudinal surveys. The sociodemographic questions included open-box questions on, e. g., listener age, gender, education, and hometown. After completing the socio-demographic questions, participants were first directed to a web page where they could adjust their headphones' volume while white noise played. They then encountered two test rounds to prepare them for the experimental part of the study. Here they listened to audio samples, which were very similar to the target guises, and were asked to rate them on a professionalism scale. This was done in order to ensure familiarity with the set-up of the study and, thus, to eradicate user errors. During the main experimental part, listeners had to give an age estimate of the speaker in form of an open-box question for each guise. The US group was then asked to rate the speaker on a professionalism scale on a discrete 6-point Likert scale. The UK group rated speakers more extensively on four discrete 6-point Likert scales: two status traits (professionalism and intelligence) as well as two social likability traits (friendliness and trustworthiness). These attributes were carefully selected following precedent (Levon and Fox 2014; Campbell-Kibler 2011b), ensuring that the results would cover both evaluative ratings of covert and overt prestige. The attributes were limited to four scales, to ensure a suitable overall length of the experiment (about twenty minutes) and consequently avoid fatigue in participants. The regional and attitudinal surveys included questions on, e. g., the way the listeners themselves speak and different ratings on how they (dis)like their own variety as well as other local varieties for the UK participants. They were also asked questions that were specifically adapted to the US and UK listener groups since both have different experiences and presumed background knowledge about Tyneside voices (see Appendix 1 and Appendix 2). Both surveys were analyzed separately.

3.2 The speakers

The study design included stimuli drawn from sociolinguistic interviews with four female and four male speakers from Tyneside who were in different age groups (see Table 1). The 20s age group consisted of speakers aged 19-24. The 40s speaker group comprised speakers aged 42–49. These age groups were chosen assuming that they capture the life stages of relatively carefree early adulthood as compared to middle adulthood, where societal pressures to conform are at their highest (see Arnett 2000, 2007; Bourdieu and Boltanski 1975; see also 2.1). All speakers were part of the "Tracing Language Variation and Change across the Life Span" (LaVaLi) corpus (Buchstaller 2024). The final selection of speakers from which stimuli were drawn were cisgender men and women. Table 1 provides an overview of the speakers using their pseudonyms and their corresponding age, gender, vernacularity score, and number of (ing) tokens per guise. The vernacularity score was determined based on a scoring system of linguistic features (Coupland 1988, 57–63, 140–156), e. g. the local realization of the FACE vowel (see 2.4) and impressionistic evaluations by the author in consultation with native speakers of English. Each speaker was given a score on the following scale: 1 = less vernacular to 5 = very vernacular.

In addition to the vernacularity score of specific speakers, vernacularity was conceptualized here via the realization of (ing). The "more vernacular" guise featured consistently [n] tokens and the "less vernacular" contained all [ŋ] tokens for all in-

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

stances of word-final (ing).[4] Instead of re-recording the speakers (see Mechler, in press), which often results in less natural speech, the guises were manipulated in PRAAT (Boersma and Weenink 2022) and Audacity to create the all [n] and the all [ŋ] guises; all other variables that occurred in the stretch of naturally occurring speech remained untouched. The following section provides more details on the guises.

Table 1: Overview over speakers categorized by age group, gender, vernacularity score (1 = "less vernacular"; 5 = "very vernacular"), and number of (ing) tokens per guise

Speaker	**Age group**	**Gender**	**Vernacularity score**	**Number of (ing) tokens**
Amy	20s	Female	3	2
Jane	20s	Female	4	1
Jason	20s	Male	2	1
Paul	20s	Male	2	1
Geraldine	40s	Female	3	1
Shannon	40s	Female	4	2
Clive	40s	Male	3	2
Jamie	40s	Male	2	3

3.3 The guises

Apart from the speakers' social properties outlined above, linguistic factors were considered in the selection of guises, which were about seven seconds long. The selection process included conversational topic to control for stylistic variation: all speakers were talking about their hobbies, and I made sure not to include pastimes that might trigger class associations such as going to the opera (with highbrow connotations). To assure comparability across samples, the selection process for guises included checking the overall coherence of the sound sample, the vernacularity of the speaker (see Table 1) and the grammatical category of the word containing the target variable (ing) (verbal / gerundial, adjective) (Schleef, Meyerhoff, and Clark 2011; Hazen 2006 inter alia). Relying on stimuli taken from sociolinguistic interviews — rather than the often used artificially created stimuli in a language lab — allowed me to collect evaluative responses based on naturally occurring speech contexts from the same corpus from the same region. This methodological choice has implications for the vernacularity score (Table 1), which describes relative differences between speakers of this sample. The overall vernacularity of the guises is high, in particular compared to guises used in previous research. For example, in his guise, Clive monophthongizes the FACE vowels in words such as way, a typical feature of Tyneside English (see 2.4). Using naturally occurring speech allows for a more nuanced understanding of situated perceptions than those typically gained in controlled experimental settings can offer. As mentioned, I will thus examine both holistic reactions and more specific responses to variants of (ing). One disadvantage of this approach is that the researcher is not

4. Some studies consider a larger number of phonetic variants for (ing), e. g. the realization of [ɪŋk], also called velar-plus-stop. Since this variant occurs in a linguistically constrained environment, mainly -thing pronominals (Clark and Asprey 2013; Mechler et al. 2022), which was not well-distributed across the spontaneous speech data used to create the stimuli in this experiment, the choice was made to restrict the tokens to a simple binary between velar and alveolar.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

able to exert as much experimental control over the stimuli. For instance, previous research has suggested that, especially for phonetic variables, several tokens are necessary for listeners to give consolidated ratings (Levon and Buchstaller 2015; Levon and Fox 2014; Wagner and Hesson 2014; Labov et al. 2011). Since the guises of the current study were not specifically created for this experiment, this was not always possible. Consequently, the final sixteen target guises featured between one and three (ing) tokens per speaker (median = 1.5) (see Table 1). Note, however, that the investigation of the random effects structure as per the regression models described below (see Table 3, 4) suggested that for status attributes listeners did not necessarily treat stimuli with one (ing) token any differently from stimuli with more than one token. For solidarity ratings, it was significant as is discussed below (see 4.3). The samples were played in randomized order, ensuring that two guises from the same speaker would not directly follow each other, to avoid listeners recognizing that two recordings were from the same speaker with velar and alveolar (ing) variants. Additionally, eight distractors from different female North-Eastern speakers of similar age, location, and gender were added. Some distractors contained variable realizations of (ing), others contained variants of FACE and GOAT; all matched the target guises in topic, genre, and length.

3.4 The listeners

To test for regional effects, data collection took place in the UK (participants: n = 49) and in the US (participants: n = 23). English listeners were predominantly from Newcastle, England, and American listeners from Columbus, Ohio, US. The majority of participants conducted the experiment on a computer with headphones in a number of quiet locations, including a museum and a language lab, as described above. Additional participants were drawn from an online version of the experiment and came from various regions in the UK. The participant sample had an age range of 18-71 years (median = 25; mean = 30.1) and was relatively balanced across gender (female: n = 42; male: n = 30). The raw data set also included two non-binary persons, who had to be excluded due to comparatively low token numbers. As is often the case in perception studies, participants were slightly skewed towards higher educational achievement (measured by highest degree attained; high school / A-levels: n = 7; college: n = 24; university: n = 40, NA: n = 1).

3.5 Data and data analysis

Ratings (professionalism: n = 1104; intelligence: n = 753; friendliness: n = 756; trustworthiness: n = 755; total: n = 3368) were elicited from altogether 72 participants. Whether participants conducted the experiment online or in person was not selected as significant factor in the model building process (see Table 3). Therefore, ratings from these groups will be collated in the following analysis. Data analysis was conducted in R (R Core Team 2024), using chi-squared analyses and Wilcoxon rank-sum tests to test significant differences between individual factors for specific samples of the data set. To reveal more complex relationships between factors, linear mixed effects regression models were built using lme4 (Bates et al. 2015) and ordinal mixed effects regression models were fit using the clmm function of the ordinal package (Christensen 2023; Freitag 2020). In a bottom-up procedure single predictors were

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

added; after the final selection of single factors, interactions were added. Model selection was informed by the Akaike information criterion (AIC), selecting models with lower AIC and significant p-values as better-fit models. Specific model builds are discussed in the relevant results sections.

4 Results

This section starts out with the results for age estimates assigned by the listeners, as we first need to establish the accuracy of speakers' perceived ages to then be able to draw conclusions of the effect of speaker age. The main part of the results section is structured according to the four speaker qualities which listeners rated: professionalism, intelligence, friendliness and trustworthiness. It focuses on the impact of these speaker qualities including (perceived) speaker age and gender since the matched-guise experiment was constructed based on the hypothesis regarding changes in the relative social costliness of speaking vernacular at different life stages (see Moelders 2025). The findings on professionalism ratings will include a section on the perception of (ing) and on the two listener locales (UK, US). The last section on friendliness ratings is used as an example to also analyze the effect of listener age and gender on speaker evaluations. Together, the results allow me to add differentiated data to ongoing work on sociolinguistic perception.

4.1 Establishing the baseline: Perceived and chronological age

One of the main questions of this study is whether speaker traits such as speaker age affect perception of their vernacularity. However, the listeners were not told how old the speakers were in real life at the time of recording. So, first, I will test how closely perceived age and real age match. In the experiment, the listeners were first asked to guess speakers' ages in an open box format. Figure 1 shows the results for these age estimates, grouped together by speaker age. Both a Wilcoxon rank-sum test ($W = 24115$, $p < .001$, $n = 1075$) and a chi-square test (χ^2 (55, $n = 1075$) $= 647.77$, $p < .001$) confirm that there is a highly significant within speaker difference. Speakers in their 20s receive lower age estimates, while middle-aged voices receive higher age estimates. Note also that the numerical spread is greater for the estimates of the four speakers in middle adulthood; the age ratings for the four younger speakers, on the other hand, seem to be more accurate. While research on the accuracy of guessing the age of voices suggests that this differential dispersion may be due to individual speaker differences (Worstbrock and Braun 2024), what is important here is that for both groups, the distribution as well as the median accurately represent the difference in speakers' real-world or chronological ages.

So, which factors mediate the age estimates? Table 2 summarizes the results of a linear mixed effects regression model fitted to the assigned age estimates (response variable) with listener age (18–29, 30–49, 50–71), listener gender (female, male), and speaker age group (20s, 40s) as fixed effects and participant and stimulus as random effects. Speaker gender, professionalism ratings, variant, and vernacularity score did not improve the model fit based on AIC calculations and were therefore dropped in the modelling process.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

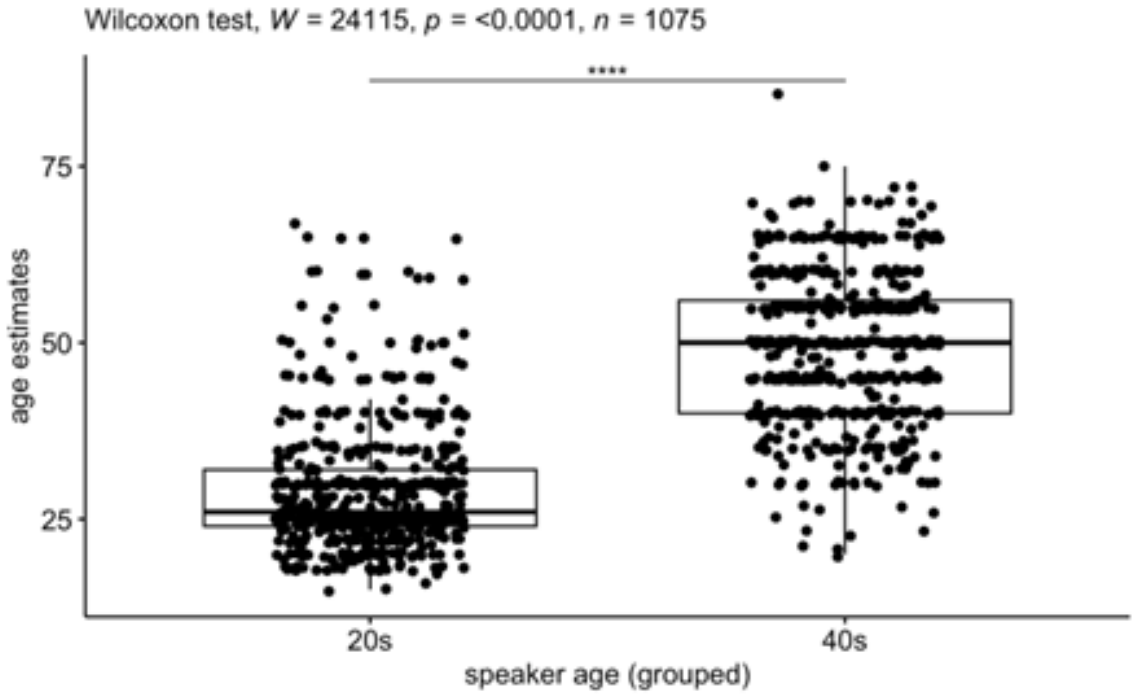

Figure 1: Age estimates (n = 1075) as assigned by listeners across two speaker age groups (20s, 40s)please provide new figures in separate files with >300 dpi

Table 2: Linear mixed effects regression model fit to age estimates (n = 1062)

Predictors (response level)	**Estimates**	**Statistic**	**p-value**	**df**
(Intercept)	33.97	16.60	**< 0.001**	1052.00
Listener age category (18–29)	- 4.42	- 2.47	**0.014**	1052.00
Listener age category (50-71)	- 7.97	- 2.91	**0.004**	1052.00
Listener gender (male)	- 6.14	- 2.52	**0.012**	1052.00
Speaker age group (40s)	20.08	10.81	**< 0.001**	1052.00
Listener age category (18–29) x listener gender (male)	3.42	1.23	0.220	1052.00
Listener age category (50–71) x listener gender (male)	11.97	3.07	**0.002**	1052.00
Random effects	**τ00**	**Std. Dev.**		
Participant (n = 71)	15.88	3.985		
Stimulus (n = 16)	12.81	3.579		
Marginal R^2 / Conditional R^2	0.529 / 0.672			

One highly significant predictor in this model is speaker age, indicating that older speakers are more likely to receive higher age estimates. In this way, the model predictions confirm previous claims based on the Wilcoxon rank-sum test and chi-square test (see above) in that the middle-aged Tyneside speakers are indeed perceived as older. Further, the model findings suggest that both listener age and gender mediate

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

age estimates: The interaction between listener age and gender predicts that male listeners aged 50 to 71 years assign significantly higher ages for the Tyneside speakers. Based on these results, we can assume that listeners (correctly) perceive younger and middle-aged speakers of Tyneside English to be of different ages. The match of perceived and chronological age confirms outcomes from previous studies which have shown that listeners are able to correctly assign a speaker's age based on their voice (Worstbrock and Braun 2024; Braun 2013). This is an important result, serving as prerequisite finding for the following analyses and it bears highlighting here as listeners were never told speakers' ages throughout the experimental setup. Rather, informants consistently judged the guises on the basis of their own age-based perception, including their assessment of how old the speakers were, which inherently triggers a range of age-associated (and gendered) stereotypes.

4.2 Evaluations of status

In the following, I present the evaluative ratings for status attributes (professionalism and intelligence) and social likability attributes (friendliness and trustworthiness) (see Montgomery and Moore 2018 inter alia).

4.2.1 Professionalism ratings

For the first status trait, the ordinal mixed effects model is fit to professionalism ratings (response variable), with speaker gender (female, male), speaker age group (20s, 40s) and place of data collection / listener locale (US, UK) as fixed effects and participant and stimulus as random effects (see Table 3).[5] Speaker age and speaker gender was also added as an interaction. The predictors variant (alveolar, velar) and vernacularity score (2, 3, 4; see Table 1) were not significant in the model building procedure (and were therefore dropped along with factors type of data collection (online, offline), number of token per guise (1, 2, 3), listener gender (female, male), listener age (18–29, 30–49, 50–71), and age estimates (numeric)).

Interestingly, ordinal regression analysis did not select variant type as significant, and the Wilcoxon rank-sum test confirms that there is only a minimal difference in the perception of the two variants ($W = 159548.5$, $p = .16$, $n = 1104$). This finding supports Levon and Fox's (2014) caution regarding the sociolinguistic salience of (ing). As in their study, my results suggest that other factors might be more meaningful for listeners in forming their judgements of the eight Tyneside speakers on the basis of naturally occurring speech. Real-time reaction experimental measures like the SLIC tool (first developed by Montgomery and Moore 2018) reveal that (ing) is in fact used as a marker for Newcastle speech (Montgomery, Vriesendorp, and Walker 2024). It is possible that in my study UK participants did pick up on the [n]–[ŋ] difference but still relied more on their overall perception of these speakers as Geordies in combination with social factors to form their judgments, not specifically (ing). So, in the following, I will focus on the (social) factors included in the model: speaker age and gender and place of data collection / listener location.

5. The threshold coefficients in Table 3 describe the likelihood of speakers being assigned to one rating category and all previous / lower categories compared to the next / higher one. For instance, Professionalism rating 3 / 4 indicates the likelihood of being rated in the rating 1 and 2 and 3 category as compared to the rating category 4.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Table 3: Ordinal mixed effects regression model fit to professionalism ratings (n = 1104; 1 = "not professional" to 6 = "very professional")

Predictors (response level)	**Odds Ratio**	**Statistic**	**p-value**	**df**
Speaker age group (40s)	1.58	1.34	0.180	Inf
Speaker gender (male)	1.31	0.80	0.422	Inf
Place of data collection (US)	0.51	- 1.70	0.088	Inf
Speaker age group (40s) × speaker gender (male)	3.80	2.75	0.006	Inf
Threshold coefficients	**Odds Ratio**	**Statistic**	**p-value**	**df**
Professionalism rating 1 / 2	0.06	- 8.27	<0.001	Inf
Professionalism rating 2 / 3	0.48	- 2.24	0.025	Inf
Professionalism rating 3 / 4	2.23	2.44	0.015	Inf
Professionalism rating 4 / 5	13.05	7.62	<0.001	Inf
Professionalism rating 5 / 6	146.05	13.17	<0.001	Inf
Random effects	**τ00**	**Std. Dev.**		
Participant (n = 72)	2.25	1.5012		
Stimulus (n = 16)	0.18	0.4284		
Marginal R^2 / Conditional R^2	**0.116 / 0.492**			

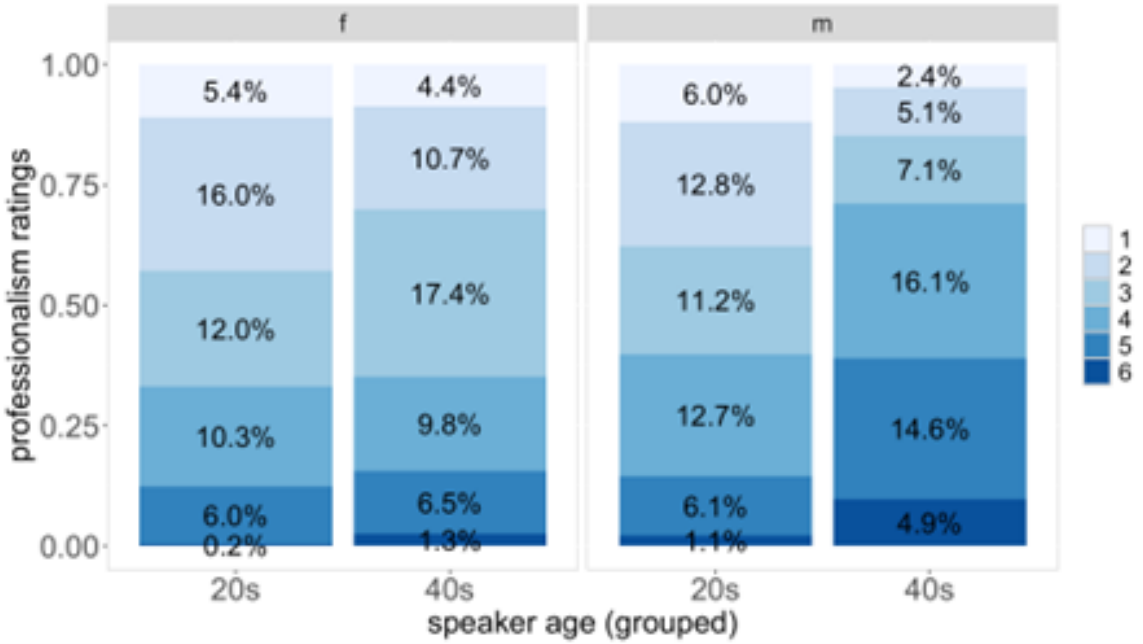

Figure 2: Professionalism ratings (n = 1104; 1 = "not professional" to 6 = "very professional") for all speakers for two speaker age groups (20s, 40s) across two speaker genders (f = female; m = male)

Figure 2 depicts professionalism ratings across speaker age and gender. Female speakers receive lower professionalism ratings than the male middle-aged adults. A

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Wilcoxon rank-sum test reveals that especially comparatively older male speakers receive significantly higher professionalism ratings than their younger counterparts ($W = 23301.5$, $p < .001$, $n = 553$). At the other end of the scale, young female individuals are rated as least professional. The prevalence of age and gender effects is also reflected in the fact that neither of the factors related to the linguistic make-up of the guises (vernacularity score and token number) was selected as significant predictor in the model building process. These results corroborate Mechler's (in press) findings that speaker age and gender play a crucial role in speaker evaluation; so much so that listeners seem to base their judgments more on these key social aspects than on the specific linguistic cues, like the variation in (ing) variants, to inform their (professionalism) ratings (see also Niedzielski 1999).

In the model for professionalism ratings, another predictor is place of data collection or listener locale (Table 3). Notably, we lack comparative data from studies with the same methodological set-up across locales to test for the locally specific indexicalities attached to different realizations of (ing) (but see Schleef and Flynn 2015 and Schleef, Flynn, and Barras 2017 for regional variation in the UK). Figure 3 plots the results for the professionalism scale across the UK and US and the two texts, whereby darker blue indicates higher professionalism ratings and lighter shades of blue mark lower professionalism ratings (note: 100% of the data are split up across both regions (50% UK, 50% US)). Adding place of data collection as fixed effect did improve the model fit, but the difference between the US and UK data was not significant (see Table 3).

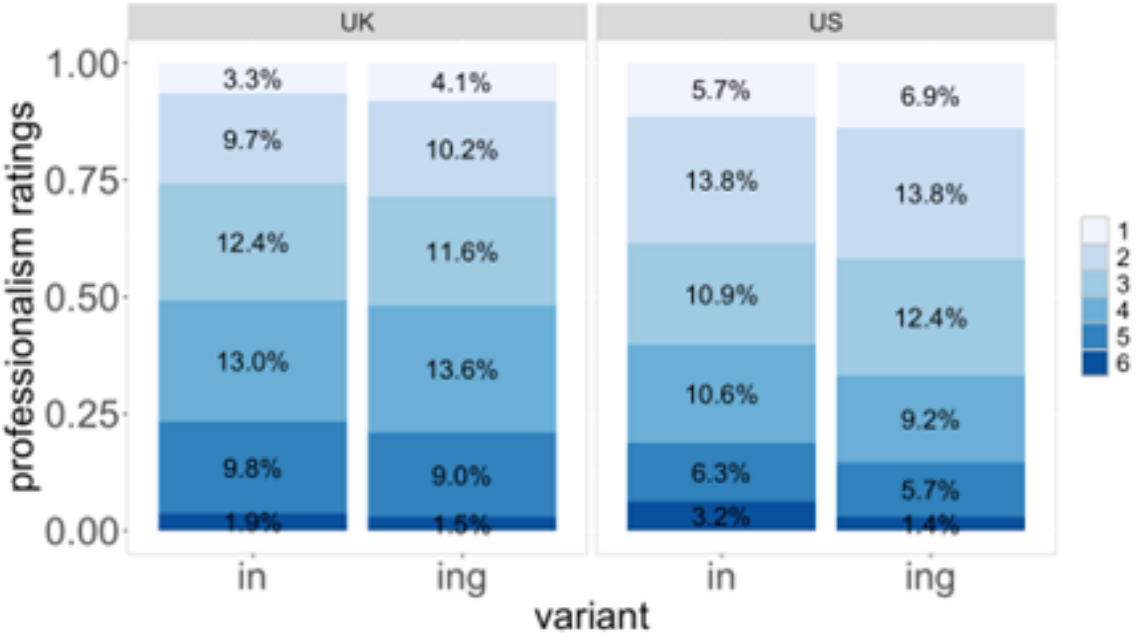

Figure 3: Professionalism ratings ($n = 1104$; 1 = "not professional" to 6 = "very professional") for all speakers for two (ing) variants ("in" = [n] and "ing" = [ŋ]) across UK and US listeners

Given that the findings in the literature claim that the alveolar realization of (ing) is highly stigmatized in the US (Labov et al. 2011; Campbell-Kibler 2008), whereas it has a much lower level of sociolinguistic salience in Northern British varieties (Schleef and

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Flynn 2015; Levon and Fox 2014), it might seem surprising at first to see US listeners rate the alveolar, more vernacular variant relatively high overall. Here the results from the language attitudes survey help us inform the interpretation of the findings (see Appendix 2, referring to statements 4 to 6): US speakers who are more confident in distinguishing different British English varieties give more nuanced ratings; listeners who are less confident assign lower ratings overall. Additionally, judgments, whereby referring to the stark socio-indexicalities tied to the (ing) variable in the US, might not be triggered for American listeners in this specific experiment because listeners do not hear a "US voice" and therefore, they do not apply the same stereotypes that they would usually associate with a US speaker. The socio-indexical value of (ing) that has typically been found for the US is not replicated then, especially when US listeners hear a "British accent" which is generally vernacular or not a very "RP sounding" voice (see Table 1). Therefore, it is not surprising to find no significant difference between the two listener locales. In sum, evaluative ratings of naturally occurring speech are predominantly mediated by speaker age and speaker gender. Importantly, phonological differences of one linguistic variable do not seem to constitute significant factors in assigning professionalism ratings, when the variable is embedded in stimuli-rich guises.

4.2.2 Intelligence ratings

Intelligence ratings, which are only available for the UK data, paint a similar picture to the results for professionalism ratings: Middle-aged male speakers are perceived as more intelligent compared to younger male speakers ($W = 10954.5$, $p < .001$, $n = 379$; see Figure 4).

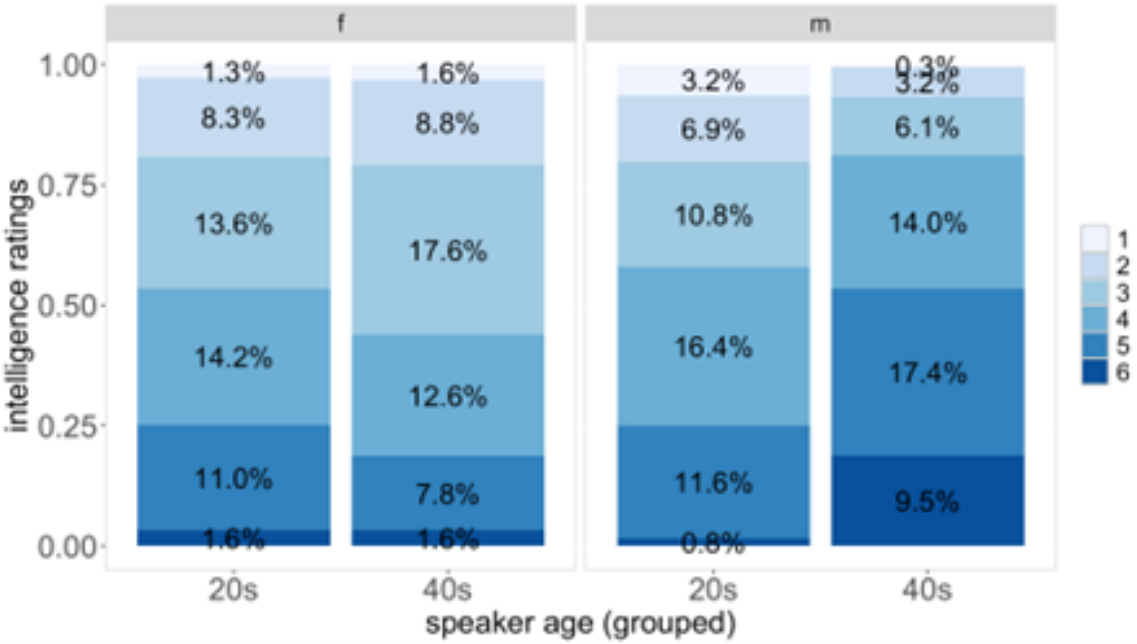

Figure 4: Intelligence ratings ($n = 753$; 1 = "not intelligent" to 6 = "very intelligent") for all speakers for two speaker age groups (20s, 40s) across two speaker genders (f = female; m = male)

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Table 4: Ordinal mixed effects regression model fit to intelligence ratings (n = 753; 1 = "not intelligent" to 6 = "very intelligent")

Predictors (response level)	**Odds Ratio**	**Statistic**	**p-value**	**df**
Speaker gender (male)	3.23	3.36	0.001	Inf
Token (2)	3.26	3.38	0.001	Inf
Token (3)	2.06	1.35	0.177	Inf
Listener gender (male)	0.40	-2.11	0.035	Inf
Threshold coefficients	**Odds Ratio**	**Statistic**	**p-value**	**df**
Professionalism rating 1 / 2	0.02	-8.57	<0.001	Inf
Professionalism rating 2 / 3	0.19	-4.08	<0.001	Inf
Professionalism rating 3 / 4	1.10	0.25	0.805	Inf
Professionalism rating 4 / 5	7.00	4.79	<0.001	Inf
Professionalism rating 5 / 6	88.21	10.09	<0.001	Inf
Random effects	**τ00**	**Std. Dev.**		
Participant (n = 72)	2.05	1.4335		
Stimulus (n = 16)	0.33	0.5724		
Marginal R^2 / Conditional R^2	**0.123 / 0.491**			

Taken together, male individuals also receive significantly higher ratings than both female age groups (W = 53344.5, p < .001, n = 753). Interestingly, there seems to be no age effect for the female speakers (W = 19125.5, p = .1, n = 374).

The results of the ordinal mixed effects regression model further replicate the gender effect (see Table 4).[6] Other factors such as speaker age, listener age, age estimates, variant, vernacularity score, and type of data collection did not improve the model fit and were thus dropped as factors. The regression results suggest that the Newcastle women are just treated as less intelligent than the men regardless of their age (note the highly significant speaker gender effect while speaker age was not selected as significant; see Table 4). I take these findings to mean that male speakers, in particular middle-aged men, have a perceptual bonus when being judged on the basis of status traits, regardless of their vernacularity (see also Mechler, in press). Hence, for both status traits ratings speaker age and/or speaker gender play a crucial role.

4.3 Evaluations of social likability

Let us now consider the results for the two social likability traits — or as I will also refer to them, solidarity traits. First, I should note that, overall, Tyneside speakers receive higher ratings in both solidarity categories than for status traits (see Figures 3, 4, 5, and 6). This is another indicator that these speakers were perceived as fairly vernacular in general (covert prestige). Crucially, when we consider the effects of speaker age and gender, the pattern is almost flipped for solidarity traits when compared to the status traits — which I will demonstrate beginning with friendliness ratings.

6. The model was fit to intelligence ratings using the following formula: `rating_intelligence ~ speaker_gender + token + listener_gender + (1|participant) + (1|stimulus)`. The effect of listener gender is explained using the example of perceived trustworthiness in section 4.3.2.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

4.3.1 Friendliness ratings

As is evident in Figure 5, younger female speakers are perceived as significantly more friendly than younger male speakers (W = 23429, p < .001, n = 377). Hence, female speakers in their 20s seem to have the "perceptual bonus" for attributes measuring social attractiveness. Interestingly, for speakers in their 40s, friendliness ratings are not mediated by gender (W = 18398.5, p = .66, n = 379).

While the results of the Wilcoxon rank-sum tests provide first indication of speaker and gender effects, when comparing specific groups, the ordinal mixed effects regression model fit to friendliness ratings gives us insight into the more complex relationship between the individual predictors.[7] Table 5 summarizes the model results for friendliness ratings. Speaker age, speaker gender, listener age, age estimates, variant, vernacularity score, and type of data collection did not improve the model fit and were thus dropped as factors. Interestingly, token number was selected as significant in this context. This is in line with previous literature, where the number of (ing) tokens impacts sociolinguistic perception (Labov et al. 2011; Levon, Buchstaller, and Mearns 2020). Here, the significance of the token number implies that the make-up of the guises has a direct effect on perceived friendliness. Another effect can be found for listener gender, where male participants assign lower ratings than female participants to the Tyneside speakers. The effects of listener traits are discussed in more detail below (see 4.3.2). In sum, friendliness ratings are meditated by speaker age and gender (based on Wilcoxon rank-sum tests); additionally, we find evidence that token number and listener gender impact perceived friendliness in naturally occurring speech.

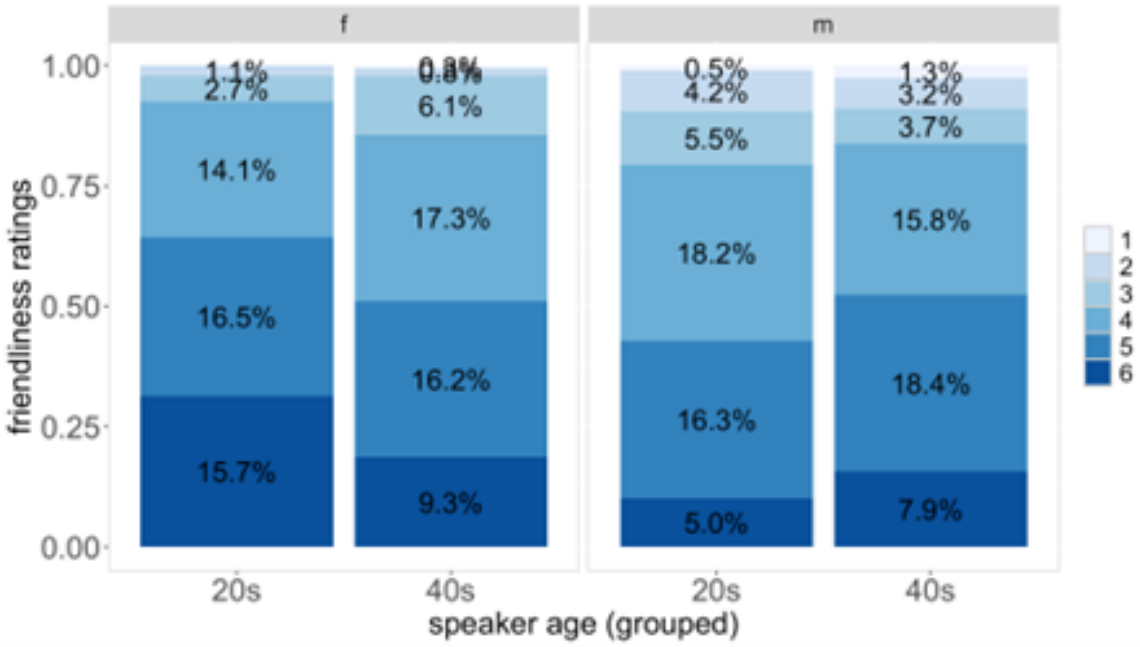

Figure 5: Friendliness ratings (n = 756; 1 = "not friendly" to 6 = "very friendly") for all speakers for two speaker age groups (20s, 40s) across two speaker genders (f = female; m = male)

7. The model was run using the following formula: `rating_friendliness ~ token + listener_gender + (1|participant) + (1|stimulus)`.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Table 5: Ordinal mixed effects regression model fit to friendliness ratings (n = 756; 1 = "not friendly" to 6 = "very friendly")

Predictors (response level)	**Odds Ratio**	**Statistic**	**p-value**	**df**
Token (2)	4.23	4.61	< 0.001	Inf
Token (3)	0.42	- 1.92	0.055	Inf
Listener gender (male)	0.25	- 3.72	< 0.001	Inf
Threshold coefficients	**Odds Ratio**	**Statistic**	**p-value**	**df**
Professionalism rating 1 / 2	0.00	- 12.01	< 0.001	Inf
Professionalism rating 2 / 3	0.02	- 11.09	< 0.001	Inf
Professionalism rating 3 / 4	0.07	- 7.99	< 0.001	Inf
Professionalism rating 4 / 5	0.71	- 1.08	0.281	Inf
Professionalism rating 5 / 6	6.66	5.88	< 0.001	Inf
Random effects	**τ00**	**Std. Dev.**		
Participant (n = 49)	1.42	1.191		
Stimulus (n = 16)	0.25	0.499		
Marginal R^2 / Conditional R^2	**0.191 / 0.463**			

4.3.2 Trustworthiness ratings

The second solidarity trait considered in this study is trustworthiness. As for friendliness ratings, speaker age and gender affect perceived trustworthiness (Figure 6). Again, female speakers in early adulthood are rated significantly more trustworthy than male speakers in their 20s (W = 23704, p < .001, n = 376) than younger male speakers. Comparing female and male speakers at a later life stage, in middle adulthood, reveals that trustworthiness for male speakers is slightly higher than for female speakers, although not significantly so (W = 17655, p = .77, n = 379).

Table 6 shows an ordinal mixed effects regression model fit to trustworthiness ratings.[8] Speaker age, speaker gender, listener age, age estimates, variant, and type of data collection did not improve the model fit and were thus dropped as factors. Similar to the results for friendliness ratings, token number affects perceived trustworthiness. But for trustworthiness, additionally speakers' vernacularity scores significantly improve the model fit. This is further confirmation that the features of naturally occurring guises have significant impact on sociolinguistic perception, especially, so it seems, on evaluations of social likability. The best-fit model for trustworthiness ratings also includes listener gender. Thus, I will discuss the impact of listener traits on the perception of the Tyneside speakers next.

8. The model was run using the following formula: `rating_trustworthiness ~ listener_gender + token + vernacularity + (1|participant) + (1|stimulus)`.

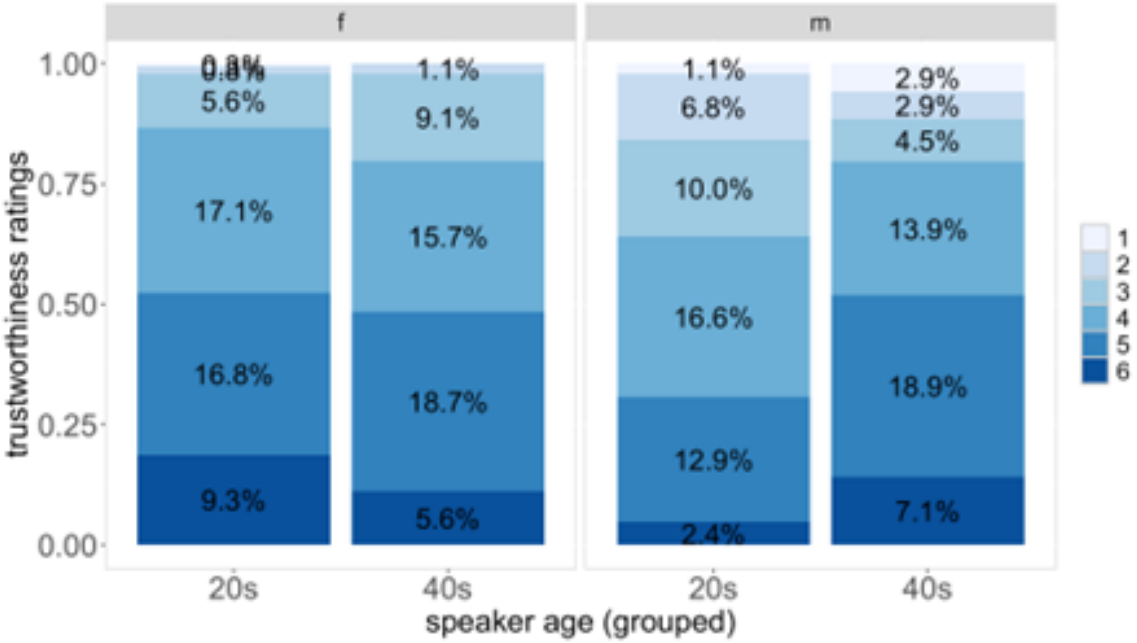

Figure 6: Trustworthiness ratings (n = 755; 1 = "not trustworthy" to 6 = "very trustworthy) for all speakers for two speaker age groups (20s, 40s) across two speaker genders (f = female; m = male)

Table 6: Ordinal mixed effects regression model fit to trustworthiness ratings (n = 755; 1 = "not trustworthy" to 6 = "very trustworthy")

Predictors (response level)	**Odds Ratio**	**Statistic**	**p-value**	**df**
Token (2)	3.45	4.79	< 0.001	Inf
Token (3)	0.89	-0.33	0.739	Inf
Speaker vernacularity score (3)	2.69	3.24	0.001	Inf
Speaker vernacularity score (4)	2.03	2.33	0.020	Inf
Listener gender (male)	0.30	-3.05	0.002	Inf
Threshold coefficients	**Odds Ratio**	**Statistic**	**p-value**	**df**
Professionalism rating 1 / 2	0.01	-11.23	< 0.001	Inf
Professionalism rating 2 / 3	0.05	-8.69	< 0.001	Inf
Professionalism rating 3 / 4	0.25	-4.21	< 0.001	Inf
Professionalism rating 4 / 5	2.15	2.38	0.017	Inf
Professionalism rating 5 / 6	28.63	9.68	< 0.001	Inf
Random effects	**τ00**	**Std. Dev.**		
Participant (n = 49)	1.61	1.2685		
Stimulus (n = 16)	0.08	0.2798		
Marginal R^2 / Conditional R^2	**0.203 / 0.473**			

Finally, then, let us consider listener-based variation in the evaluation of Tyneside voices in more detail. Figure 7 depicts friendliness ratings (n = 756) for all speakers combined across three listener age groups and two genders. Firstly, it exemplifies a

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

pattern that holds true for perceived intelligence (Table 4), friendliness (Table 5), and trustworthiness (Table 6): a highly significant gender effect for listeners, whereby male participants assign considerably lower ratings (e. g., Levon et al. 2021). Listeners in their 50s to 70s evaluating speakers more harshly could be indicative of an increased critical position listeners take towards speaker behaviors considered inappropriate in certain contexts as they get older. Levon et al. (2021) report on an age-split around 45 years, with older respondents behaving more judgmental in the surveys than younger ones. This age-split is corroborated by the above findings as (male) listeners above 40 clearly evaluate speakers differently than respondents in their 30s and younger. Although listener age was not significant in the regression analysis (see Table 6), the distributional results still highlight a trend: As people age, they become increasingly critical of the speech patterns of people younger than themselves — possibly a facet of the "complaint tradition" (see Buchstaller and Traugott 2006).

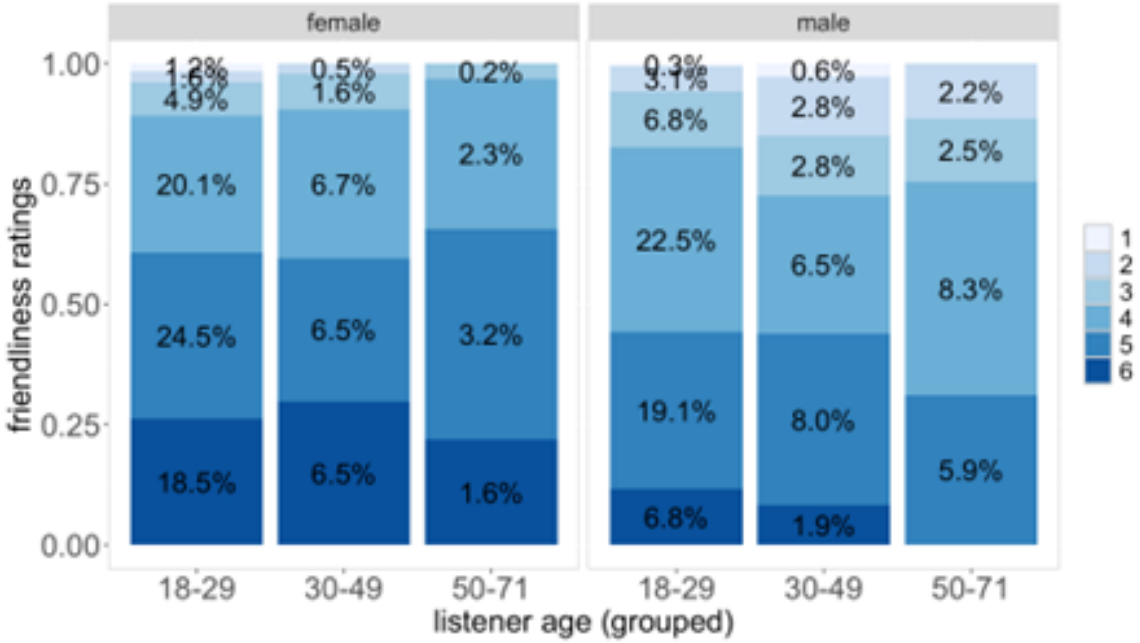

Figure 7: Friendliness ratings (n = 756; 1 = "not friendly" to 6 = "very friendly") for all speakers across listener age and gender

Overall, vernacular voices do affect perceived professionalism, intelligence, friendliness, and trustworthiness, but in different ways (RQ1). Stimuli-rich voice samples were rated as overall more friendly and trustworthy than professional and intelligent (covert prestige). Speaker age and gender were found to be relevant predictors for all categories, but most importantly for status traits (where the factors were significant in the regression models) (RQ2). Lastly, for all attributes certain participant traits were selected in the final model: for professionalism listener region, for intelligence, friendliness, and for trustworthiness listener gender (and age) (RQ3).

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

5 Discussion

The aim of this study was to extend the framework of sociolinguistic research by conducting a matched-guise experiment with naturally occurring, stimuli-rich guises from sociolinguistic interviews and to test effects on status and solidarity traits (RQ1). Between the two (ing) variants, I could only find minimal perceptual differences. This result might seem unexpected at first given previous research but can be explained by the nature of the guises used in the experiment. First, this study relies on samples from semi-natural speech, whereas most other studies focus on speech in highly controlled environments (e. g., Labov et al. 2011). It is important to remember that this study included one to three (ing) tokens per speaker — previous research suggests that listeners might simply need more occurrences of (ing) for consolidated ratings (Levon and Fox 2014; but see Mackay, Schleef, and Pflaeging 2024). For example, Labov et al. (2011) show that a proportion of 30% apical variants is needed to elicit an effect. Thus, the methodological set-up could have had an impact on listeners judgments as they are potentially less aware of the (ing) variants (Mackay, Schleef, and Pflaeging 2024), leading to more holistic evaluations (Vaughn 2022). Further, it might be the case that this phenomenon applies predominantly to shorter guises, which contain fewer linguistic cues and thus less information about the speakers in the actual speech stream, which might possibly predispose raters to draw on pre-existing stereotypes. For instance, while Vaughn and Kendall's 2018 study on listeners' sensitivity to variants of (ing) implies "that listeners can and do form probabilistic expectations about the realization of (ING)" (Vaughn and Kendall 2018, 69), this effect might not apply to naturally occurring guises. Subsequently, my study might speak less to the socio-indexicalities of (ing) specifically, but more to the overall perception of Tyneside speakers and common stereotypes based on speakers' and listeners' age and gender. Note here, that my findings are fully in line with findings on attitudes towards North-Eastern varieties in that the Tyneside voices generally receive higher social likability ratings than status ratings (Levon et al. 2021 inter alia).

What is the social costliness of speaking vernacular at different life stages, then (RQ2)? Listeners rate male speakers in their 40s higher on professionalism and intelligence than female speakers, even when they consistently produce the vernacular [n] variant and irrespective of their overall relatively high vernacularity score. Younger female speakers are perceived as significantly more friendly and trustworthy than younger male speakers, again irrespective of their vernacularity. These results underscore the findings from previous studies that gender impacts the social indexicality of linguistic variables (Cole 2021; Boyd, Fruehwald, and Hall-Lew 2021 inter alia). Following Niedzielski's 1999 claim that social factors might override acoustic information, I argue that (the features associated with) perceived speaker gender and age are overriding the specific acoustic information carried by (ing) and/or vernacularity. By extension, listener reactions reflect the stereotypes and biases of hegemonic structures in society, including the attribution of status traits to men, especially middle-aged men, and of social likability traits to women.

An additional factor to consider is that of listener traits (RQ3). The almost identical study design across the two locales allowed for direct comparison of evaluative ratings, and the analysis showed some differences between professionalism ratings for US and UK listeners (see Table 3). The contrast was minimal, but, as pointed out, this may be due to aspects of the stimulus design or the fact that US listeners

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

were rating a UK accent (and an unfamiliar one at that). Regarding UK listeners, Levon and Fox (2014) also did not find an effect for professionalism ratings, so my findings add further evidence to the regionally varied perception of (ing). As outlined above, listener age and gender constitute an important factor in perception, but we lack detail on the fine-grained process via which age- and gender-related stereotypes affect speaker evaluation. The present study can contribute insights into the effect of listener characteristics in perception research (see also Cole 2021). The findings corroborate previous studies in that older male listeners are more critical than their younger counterparts (e. g., Levon et al. 2021). One explanation for this finding is that older listeners might have been longer exposed to the socio-normative pressures of the adult marketplace and internalized them (see 2.2), which in turn would lead to them being more judgmental towards variability that is deemed inappropriate in more monitored contexts (see Labov et al. 2011 for a similar argument). An alternative explanation relies on a generational effect, in that there is change towards more tolerance towards vernacular forms. This hypothesis is supported by recent evidence from socio-perceptual studies in the workplace: Levon et al. (2021) report on a cohort effect towards higher acceptance of dialects in the professional environments.

6 Conclusion

This study set out to test the impact of different life stages on speaker evaluation by comparing perceptual ratings of speakers in early and middle adulthood. In sum, the results show that evaluations are highly contingent on speaker age and gender as well as listener age and gender. These findings have methodological and theoretical implications. Methodologically, this perceptual experiment underlines the importance of taking more seriously than previously the impact of deeply rooted social stereotypes when working with naturally occurring, stimuli-rich guises. In this sense, they provide an age-and gender-differentiated continuation of Niedzielski (1999)'s classic work which has drawn attention to importance of social as compared to linguistic cues in perception research. In future work, the use of visuals may help carefully control for these stereotypes (Campbell-Kibler 2021). On the level of theory, the results contribute to both age-grading theory and the study of sociolinguistic perception: First, they draw our attention to the interactional nature of age and gender in sociolinguistic perception (see also Levon et al. 2021), providing a finely differentiated picture of the ways in which age- and gender-related stereotypes impact speaker evaluation (Cole 2021). The gender and age effects highlight the importance of social variables as possible motivations for age-grading of (ing) as a phenomenon in speech production. And similarly, while the variable under scrutiny, (ing), was not very relevant to how listeners evaluated speakers, especially in the context of spontaneous speech, my research provides new insights into how social perceptions might prompt differences in individual linguistic choices across the lifespan. Hence, overall, the findings reported here allow us to develop a better understanding of the relationship between the perception of linguistic choices and how these patterns might be related to the socio-indexical meaning of voices of all ages.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Funding

Project "Tracing language variation and change across the life span" (LaVaLi), funded by Deutsche Forschungsgemeinschaft (DFG), Project number 418602718, PI Isabelle Buchstaller, 2019–2024, https://gepris.dfg.de/gepris/projekt/418602718.

7 Ethics

Official approval to conduct this study was obtained from the ethics committee for humanities at the University of Duisburg-Essen, Germany. Contact information: Prof. Dr. Isabelle Buchstaller, Universitätsstr. 12, 45117 Essen, i.buchstaller@uni-due.de.

For data collection in the US, official approval was also granted by the Institutional Review Board at The Ohio State University, USA. Contact information: Enterprise for Research, Innovation and Knowledge, University Square South, 15 E. 15th Avenue, 4th Floor, Columbus OH 43201, US, IRBinfo@osu.edu.

Conflict of interest

The authors have no conflict of interest to declare.

References

Arnett, Jeffrey J. 2000. "Emerging adulthood: A theory of development from the late teens through the twenties." *American Psychologist* 55 (5): 469–480. https://doi.org/10.1037//0003-066X.55.5.469.

———. 2007. "Emerging adulthood: What is it, and what is it good for?" *Child Development Perspectives* 1:68–73. https://doi.org/10.1111/j.1750-8606.2007.00016.x.

Ball, Peter. 1983. "Stereotypes of Anglo-Saxon and non-Anglo-Saxon accents: Some explanatory Australian studies with the matched-guise technique." *Language Sciences* 5:163–184. https://doi.org/10.1016/S0388-0001(83)80021-7.

Bates, Douglass, Martin Mächler, Ben Bolker, and 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.

Beal, Joan C., Lourdes Burbano-Elizondo, and Carmen Llamas. 2012. *Urban North-Eastern English: Tyneside to Teeside.* Edinburgh: Edinburgh University Press.

Beaman, Karen. 2021. "Coherence in real- and apparent-time: A sociolinguistic variationist investigation of language change in Swabia." Doctoral dissertation, Queen Mary University of London.

Boersma, Paul, and David Weenink. 2022. *Praat: Doing phonetics by computer (computer program).* Version 6.3.03, retrieved 17 December 2022. https://www.praat.org/.

Bourdieu, Pierre, and Luc Boltanski. 1975. "Le fétichisme de la langue." *Actes de la recherche en sciences sociales* 4:2–32.

Boyd, Zac, Josef Fruehwald, and Lauren Hall-Lew. 2021. "Crosslinguistic perceptions of /s/ among English, French, and German listeners." *Language Variation and Change* 33 (2): 165–191. https://doi.org/10.1017/S0954394521000089.

Braun, Angelika. 2013. "Age estimation by different listener groups." *International Journal of Speech, Language and the Law* 3:65–73. https://doi.org/10.1558/ijsll.v3i1.65.

Buchstaller, Isabelle. 2024. "Tracing Language Variation and Change across the Lifespan (LaVaLi)." https://gepris.dfg.de/gepris/projekt/418602718.

Buchstaller, Isabelle, Anne Krause, Anja Auer, and Stefanie Otte. 2017. "Levelling across the life-span?: Tracing the FACE vowel in panel data from the North East of England." *Journal of Sociolinguistics* 21:3–33. https://doi.org/10.1111/josl.12227.

Buchstaller, Isabelle, and Elizabeth Closs Traugott. 2006. "*The lady was al demonyak*: Historical aspects of adverb *all*." *English Language and Linguistics* 10 (2): 345–370. https://doi.org/10.1017/S136067430600195X.

Campbell-Kibler, Kathryn. 2007. "Accent, (ING), and the social logic of listener perceptions." *American Speech* 82 (1): 32–64. https://doi.org/10.1215/00031283-2007-002.

———. 2008. "I'll be the judge of that: Diversity in social perceptions of (ING)." *Language in Society* 37 (5): 637–659. https://doi.org/10.1017/S0047404508080974.

———. 2009. "The nature of sociolinguistic perception." *Language Variation and Change* 21 (1): 135–156. https://doi.org/10.1017/S0954394509000052.

———. 2011a. "Intersecting variables and perceived sexual orientation in men." *American Speech* 86 (1): 52–68. https://doi.org/10.1215/00031283-1277510.

———. 2011b. "The sociolinguistic variant as a carrier of social meaning." *Language Variation and Change* 22:423–441. https://doi.org/10.1017/S0954394510000177.

———. 2021. "Deliberative control in audiovisual sociolinguistic perception." *Journal of Sociolinguistics* 25 (2): 253–271. https://doi.org/10.1111/josl.12418.

Childs, Claire. 2013. "'I couldn't really put [mə] finger on it': Phonetic realisations of the possessive singular 'my' in Tyneside English." *Newcastle Working Papers in Linguistics: Selected Papers from Sociolinguistics Summer School 4* 19 (1): 42–66.

Christensen, Rune H. B. 2023. *ordinal: Regression models for ordinal data.* R package version 2023.12-4, https://CRAN.R-project.org/package=ordinal.

Clark, Urszula, and Esther Asprey. 2013. *West Midlands English: Birmingham and the Black Country.* Edinburgh: Edinburgh University Press.

Cole, Amanda. 2021. "Disambiguating language attitudes held towards socio-demographic groups and geographic areas in South East England." *Journal of Linguistic Geography* 9 (1): 13–27. https://doi.org/10.1017/jlg.2021.2.

Coupland, Nikolas. 1988. *Dialect in use: Sociolinguistic variation in Cardiff English.* Cardiff: University of Wales Press.

Coupland, Nikolas, and 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.

Downes, William. 1998. *Language and society.* 2nd. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139163781.

Eckert, Penelope. 1997. "Age as a sociolinguistic variable." In *The handbook of sociolinguistics,* edited by Florian Coulmas, 151–167. Oxford: Blackwell.

Fehringer, Carol, and Karen P. Corrigan. 2015. "The rise of the going to future in Tyneside English: Evidence for further grammaticalisation." *English World-Wide* 36 (2): 198–227. https://doi.org/10.1075/eww.36.2.03feh.

Forrest, Jon, and Walt Wolfram. 2019. "The status of (ing) in African American Language: A quantitative analysis of social factors and internal constraints." *American Speech* 94 (1): 72–90. https://doi.org/10.1215/00031283-7308049.

Freitag, Raquel Meister Ko. 2020. "Effects of the linguistic processing: Palatals in Brazilian Portuguese and the sociolinguistic monitor." *University of Pennsylvania Working Papers in Linguistics* 25 (2): 21–30.

Giles, Howard. 1970. "Evaluative reactions to accents." *Educational Review* 22:211–227. https://doi.org/10.1080/0013191700220301.

Grama, James, Isabelle Buchstaller, Anne-Marie Moelders, Lea Bauernfeind, and Mirjam E. Eiswirth. in press. "Ageing in style: Towards disentangling style-shifting and lifespan change." In *Connecting the individual and the community: Contributions from sociolinguistic panel research,* edited by Karen V. Beaman and Isabelle Buchstaller. Routledge.

Grama, James, Johanna Mechler, Lea Bauernfeind, Mirjam E. Eiswirth, and Isabelle Buchstaller. 2023. "Post-educator relaxation in the U-shaped curve: Evidence from a panel study of Tyneside (ing)." *Language Variation and Change* 35 (3): 325–350. https://doi.org/10.1017/S0954394523000194.

Guimarães, Thiago. 2014. "Revealed: The demographic trends for every social network." *Business Insider Career* September 3. https://www.businessinsider.com/2014-social-media-demographics-update-2014-7.

Hazen, Kirk. 2006. "IN/ING Variable." In *Encyclopedia of language and linguistics,* 2nd, edited by Keith Brown, 5:581–584. Oxford: Elsevier. https://doi.org/10.1016/B0-08-044854-2/04716-7.

Labov, William. 2001. *Principles of linguistic change: Volume 2: Social factors.* Oxford: Blackwell.

Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski, and Naomi Nagy. 2011. "Properties of the sociolinguistic monitor." *Journal of Sociolinguistics* 15:431–463. https://doi.org/10.1111/j.1467-9841.2011.00504.x.

Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner, and Samuel Fillenbaum. 1960. "Evaluational reactions to spoken languages." *Journal of Abnormal and Social Psychology* 60:44–51. https://doi.org/10.1037/h0044430.

Lavan, Nadine, and Carolyn McGettigan. 2023. "A model for person perception from familiar and unfamiliar voices." *Communications Psychology* 1 (1): 1–11. https://doi.org/10.1038/s44271-023-00001-4.

Levon, Erez, and Isabelle Buchstaller. 2015. "Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing." *Language Variation and Change* 27 (3): 319–348. https://doi.org/10.1017/S0954394515000149.

Levon, Erez, Isabelle Buchstaller, and Adam Mearns. 2020. "Towards an integrated model of perception: Linguistic architecture and the dynamics of sociolinguistic cognition." In *Advancing socio-grammatical variation and change: In honour of Jenny Cheshire,* edited by Karen V. Beaman, Isabelle Buchstaller, Susan Fox, and James A. Walker, 32–54. London: Routledge.

Levon, Erez, and Sue Fox. 2014. "Social salience and the sociolinguistic monitor: A case study of ING and TH-fronting in Britain." *Journal of English Linguistics* 42:185–217. https://doi.org/10.1177/0075424214531487.

Levon, Erez, Devyani Sharma, Dominic J. L. Watt, Amanda Cardoso, and Yang Ye. 2021. "Accent bias and perceptions of professional competence in England." *Journal of English Linguistics* 49 (4): 355–388. https://doi.org/10.1177/00754242211046316.

Mackay, Bradley, Erik Schleef, and Jana Pflaeging. 2024. "Awareness and sociolinguistic monitoring: Comparing (ing), (t)-deletion, Like, You Know and Pauses." In *ICLaVE12.* University of Vienna.

McKenzie, Robert M., and Andrew McNeill. 2022. *Implicit and explicit language attitudes: Mapping linguistic prejudice and attitude change in England.* New York: Routledge. https://doi.org/10.1080/01434632.2023.2250959.

Mechler, Johanna. in press. "Investigating age effects in the perception of (ing): A study on professionalism ratings from the North East of England." In *Connecting the individual and the community: Contributions from sociolinguistic panel research,* edited by Karen V. Beaman and Isabelle Buchstaller. Routledge.

Mechler, Johanna, and Isabelle Buchstaller. 2019. "[In]stability in the use of a stable variable." *Linguistic Vanguard* 5:1–13. https://doi.org/10.1515/lingvan-2018-0024.

Mechler, Johanna, James Grama, Lea Bauernfeind, Mirjam Eiswirth, and Isabelle Buchstaller. 2022. "Towards an empirically-based model of age-graded behaviour: Trac(ing) linguistic malleability across the entire adult life-span." *University of Pennsylvania Working Papers in Linguistics* 28 (2): 101–110.

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

Montepare, Joann M., Daniel Kempler, and Tracy McLaughlin-Volpe. 2014. "The voice of wisdom: New insights on social impressions of aging voices." *Journal of Language and Social Psychology* 33 (3): 241–259. https://doi.org/10.1177/0261927X13519080.

Montgomery, Chris. 2007. "Northern English dialects: A perceptual approach." Doctoral dissertation, University of Sheffield.

Montgomery, Chris, and Emma Moore. 2018. "Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech." *Language* 94 (3): 629–661. https://doi.org/10.1353/lan.2018.0038.

Montgomery, Chris, Hielke Vriesendorp, and Gareth Walker. 2024. "Real-time accent recognition of five Northern Englishes. Do listeners (still) distinguish between Tyneside, Merseyside, Manchester, Sheffield and Leeds English?" In *NEW10.* Newcastle University.

Niedzielski, Nancy. 1999. "The effect of social information on the perception of sociolinguistic variables." *Journal of Social Psychology* 18 (1): 62–85. https://doi.org/10.1177/0261927X99018001005.

Pfenninger, Simone E. 2024. "Language learning in later life: From significant life events to cognitive benefits." In *CLARe6.* University of Tübingen.

R Core Team. 2024. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Https://www.R-project.org. Vienna, Austria.

Schleef, Erik, and Nicholas Flynn. 2015. "Ageing meanings of (ing) in Manchester, England." *English World-Wide* 36:48–90. https://doi.org/10.1075/eww.36.1.04sch.

Schleef, Erik, Nicholas Flynn, and Will Barras. 2017. "Regional diversity in social perceptions of (ing)." *Language Variation and Change* 29:29–56. https://doi.org/10.1017/S0954394517000047.

Schleef, Erik, Miriam Meyerhoff, and Lynn Clark. 2011. "Teenagers' acquisition of variation: A comparison of locally-born and migrant teens' realisation of English (ing) in Edinburgh and London." *English World-Wide* 32:206–236. https://doi.org/10.1075/eww.32.2.04sch.

Schmitt, Antje, and Dana Unger. 2019. "Lifespan perspective on occupational health." In *Work Across the Lifespan,* edited by Boris B. Baltes, Cort W. Rudolph, and Hannes Zacher, 369–393. Elsevier. https://doi.org/10.1016/B978-0-12-812756-8.00015-3.

Stewart, Mark A., and Ellen Bouchard Ryan. 1982. "Attitudes toward younger and older adult speakers: Effects of varying speech rates." *Journal of Language and Social Psychology* 1 (2): 91–109. https://doi.org/10.1177/0261927X8200100201.

Tamminga, Meredith. 2019. "Sources of microtemporal clustering in sociolinguistic sequences." *Frontiers in Artificial Intelligence* 2 (10): 1–17. https://doi.org/10.3389/frai.2019.00010.

Travis, Catherine E., James Grama, and Benjamin Purser. 2023. "Stability and change in (ing): Ethnic and grammatical variation over time in Australian English." *English World-Wide* 44 (3): 435–469. https://doi.org/10.1075/eww.22043.tra.

JLAR 3 (2025)
10.15460/jlar .2025.3.1.XXX

Trudgill, Peter. 1974. *The social differentiation of English in Norwich.* Cambridge: Cambridge University Press.

Vall, Natasha. 2007. *Cities in decline?: A comparative history of Malmö and Newcastle after 1945.* PLACE: Malmö University.

Vaughn, Charlotte. 2022. "What carries greater social weight, a linguistic variant's local use or its typical use?" In *Proceedings of the 44th Annual Conference of the Cognitive Science Society,* edited by Jennifer Culbertson, Andrew Perfors, Hugh Rabagliati, and Veronica Ramenzoni, 44:512–518.

Vaughn, Charlotte, and Tyler Kendall. 2018. "Listener sensitivity to probabilistic conditioning of sociolinguistic variables: The case of (ING)." *Journal of Memory and Language* 103:58–73. https://doi.org/10.1016/j.jml.2018.07.006.

Wagner, Suzanne Evans. 2012a. "Age grading in sociolinguistic theory." *Language and Linguistics Compass* 6:371–382. https://doi.org/10.1002/lnc3.343.

———. 2012b. "Real-time evidence for age grad(ing) in late adolescence." *Language Variation and Change* 24:179–202. https://doi.org/10.1017/S0954394512000099.

Wagner, Suzanne Evans, and Ashley Hesson. 2014. "Individual sensitivity to the frequency of socially meaningful linguistic cues affects language attitudes." *Journal of Language and Social Psychology* 33 (6): 651–666. https://doi.org/10.1177/0261927X14528713.

Watt, Dominic. 2000. "Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated?" *Language Variation and Change* 12:69–101. https://doi.org/10.1017/S0954394500121040.

———. 2002. "'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system." *Journal of Sociolinguistics* 6 (1): 44–63. https://doi.org/10.1111/1467-9481.00176.

Watt, Dominic, and Lesley Milroy. 1999. "Patterns of variation and change in three Newcastle Vowels: Is this dialect levelling?" In *Urban voices: Accent studies in the British Isles,* edited by Paul Foulkes and Gerard Docherty, 25–46. London: Routledge.

Wells, John C. 1982. *Accents of English 2: The British Isles.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511611759.

Worstbrock, Carolin, and Angelika Braun. 2024. "Estimating speaker age from voices." In *CLARe6.* University of Tübingen.