Local versus Global Cues for a Speaker to be Disfluent: Feeling of Another's Knowing for Native and Non-native Speakers

Esperanza Badaya[1*], Robert Hartsuiker[1], and Martin Corley[2]

[1]Department of Experimental Psychology, Ghent University, Ghent, Belgium.

[2]Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom.

[*]Corresponding author. Email: esperanza.badaya@ugent.be

Abstract

Filled pauses, such as *uh* and *um* in English, are commonly attributed by listeners to a given speaker's lack of confidence in their knowledge on a particular topic (Feeling of Another's Knowing, Brennan & Williams, 1995). However, speakers might be hesitant for reasons other than confidence: For example, non-native-accented speech contains more hesitation markers than its native counterpart (Davis, 2003; De Jong, 2016). Previous literature has shown that listeners are indeed sensitive to differences in filled pause rates as a function of speakers' linguistic backgrounds (e.g., Bosker et al., 2014; Bosker et al., 2019). A potential account of this effect concerns non-native speakers' stereotyped lower linguistic competence (e.g., Lev-Ari, 2015). Consequently, in this experiment, we set up to explore whether this lower linguistic competence extends to non-linguistic elements, such as filled pauses, via a *horse race task*, in which we indirectly measured listeners' feeling of another's knowing as reflected in the amount of money bet on a series of horses following a native and a non-native speakers descriptions.

1 Introduction

In situations when the facts are not known, we usually need to decide whether to believe what we are being told, either by considering the content of the communicated information or by its source. Individuals are equipped with a set of cognitive abilities to monitor whether they are being misinformed (i.e., epistemic vigilance, Sperber et al., 2010). The trustworthiness[1] of a speaker can be thought to be assessed against the perception of them having genuine information and their intention to share it with their interlocutor (Sperber et al., 2010): i.e., they are competent and they are honest. Assuming that a speaker is honest, attributions of competence can be biased by several, potentially irrelevant, factors, ranging from the situational (particularities about the speaker) to the ephemeral (particularities about the details of speech). Here, we explore whether the weight assigned to disfluency (i.e., interruptions to the flow of speech that do not add propositional content, Fox Tree, 1995), a paralinguistic cue previously reported to affect a listener's perception of a given speaker's confidence in their knowledge, can be diminished when speakers have alternative reasons to be disfluent: e.g., producing speech in their second language. In what follows, we briefly review the evidence showing the separate effects of speaker identity and manner of delivery on trust, before discussing their combined effects. It is important to note that of interest for us is not whether listeners believe the speaker to be intentionally misleading (i.e., they are trying to deceive them); rather, we are interested in whether, in a situation where the speaker is expected to be cooperative and knowledgeable, listeners' decisions on trust are still biased by irrelevant cues.

Individuals display biases against speakers which are triggered by social information, even when these biases are unfounded. One social category that can impact listeners' evaluations is a speaker's linguistic background—that is, whether the speaker is producing speech in their first or second language. A potential explanation for this bias has its origins in the ease with which non-native-accented speech is processed. Lev-Ari and Keysar (2010) had participants listen to trivia statements read out loud by either a native or a non-native English speaker. Native English listeners were asked to rate the veracity of the statements. Across the board, non-native-produced utterances were rated as less truthful, even when participants were made aware of the biases associated with a speaker's accent (Exp. 2). As participants had been told that the trivia statements were read aloud, these findings suggest that they weren't making judgements about each speaker's knowledge, but rather, basing their evaluations on the difficulties associated with processing accented speech (i.e., processing fluency, see also Dragojevic, Giles, Beck, & Tatum, 2017). Although Lev-Ari and Keysar's (2010) behavioural findings have not been widely replicated (Foucart & Hartsuiker, 2021), and both stereotyping and processing fluency could lead to the biases previously reported towards

[1] Trustworthiness has also been referred to as *beliavility* or *credibility* elsewhere. We use them here interchangeably to refer to listeners' believing what they are told.

non-natives (Mai & Hoffmann, 2013), their results align with a wider field of research where speaker identity not only affects how speech is processed but how it is interpreted, and in particular, with research highlighting differences in processing when speech is produced with a native or a non-native accent (e.g., Hanulíková, van Alphen, van Goch, & Weber, 2012; Bosker, Quené, Sanders, & De Jong, 2014; Foucart, Costa, Morís-Fernández, & Hartsuiker, 2020).

The details (as opposed to the content) of speech can also affect the determination of trust. For example, vocally expressed confidence via prosody affects listeners' judgments of believability (Jiang & Pell, 2017); and even when foreign accents are less trusted, this can also be modulated by how confident their voice sounds (Jiang, Gossack-Keenan, & Pell, 2019; Caballero & Pell, 2020). Importantly to our aims, Brennan and Williams (1995) demonstrated that filled pauses are perceived as an index of speakers' confidence in their own knowledge. In their study, participants listened to previously recorded answers to trivia questions (without hearing the questions) and were asked to rate how likely each speaker would be to recognise the correct answer to the question: What the authors termed, from the listener's perspective, *Feeling Of Another's Knowing* (FOAK). Amongst the cues that biased participants' assessments were filled pauses: Answers riddled with disfluencies were more likely to receive lower FOAK ratings, suggesting that filled pauses were taken as reflective of speakers' reduced certainty about their knowledge; in contrast, non-answers (i.e., 'I don't know') were more likely to receive a higher FOAK rating if preceded by a filled pause. Brennan and Williams took this as evidence that listeners are sensitive to the surface form of delivery, and in particular, to the cues displayed by speakers when they do not know (or cannot remember at the moment of being asked) the answer to a question. One consequence of a perception that speakers have reduced confidence in what they are saying could be that, in return, listeners place less trust in them (see Jiang et al. 2017).

A speaker can however be disfluent for reasons other than reduced certainty in their knowledge. For example, producing speech in one's second language can also lead to disfluent speech. Speaking a second language is cognitively demanding (Gregersen, 2005) and many of the difficulties associated with language production in second language (e.g., word-finding difficulties, Pivneva, Palmer, & Titone, 2012) can be associated with disfluency (e.g., Beattie & Butterworth, 1979; Schachter, Christenfeld, Ravina, & Bilous, 1991). Recently, Matzinger, Pleyer, and Żywiczyński (2023) explored whether listeners' perceptions of why a speaker was disfluent differed for native and non-native speakers. In their study, participants listened to native and non-native speakers answer trivial questions (i.e., FOAK) and answer requests (i.e., willingness perceptions), and were explicitly asked to rate the speaker's knowledge and confidence (for knowledge perceptions) and their willingness to grant the request. Crucially, Matzinger et al. (2023) manipulated speakers' fluency by having answers prefaced with either short (200 ms) or long (1200 ms) pauses. While long pauses produced by a non-native speaker were less likely to be rated as reflecting unwillingness on the speaker's behalf to grant a request in contrast to natives' long pauses, FOAK ratings did not differ between speakers: Long pauses produced by either speaker were likely to be taken as reflecting low confidence and low knowledge. Matzinger et al. (2023) argue this striking finding could be explained due to the different conversational contexts: Requests tap onto speakers' cooperativeness, and thus tunning to the interlocutor's mental state might be more relevant in contrast to evaluating the speaker's competence (i.e., knowledge).

Most of the studies herein described have asked listeners explicitly what dimension to assess. In these studies, participants are commonly asked to judge on a scale how confident, trustworthy, or knowledgeable the speaker sounds, and authors have concluded how different details of speech and of speakers impact listeners' evaluations. There is ample evidence suggesting that individuals' behaviours differ depending on task demands and, specifically, there are differences between individuals' responses when they are explicitly asked to judge a dimension and their responses via implicit measurements. In the case of disfluencies in native and non-native speech, it could be possible that individuals take them differently in the face of assessing knowledge when there is no explicit evaluation of the speaker. For example, in contexts where listeners' task is just to follow instructions, filled pauses produced by a non-native speaker do not lead to anticipatory eye-movements towards elements with low-frequency labels, in contrast to disfluencies produced by a native speaker (Bosker, Quené, Sanders, & De Jong, 2014). Similarly, individuals' behaviours upon encountering native and non-native (dis)fluent speech might be different, even when they consciously attribute disfluency to the same reasons.

1.1 The present study

To explore this idea, we propose an experiment wherein participants' behaviour depends on their assessment of the speaker's knowledge with a new paradigm: The horse-race paradigm. In this task, participants are presented with one (or more) speaker, said to be well-known tipsters, who will provide information about four horses who will run in an upcoming race. In turn, participants are asked to place bets on each horse given what they are told. Crucially, information can be delivered fluently or disfluently. This paradigm presents two advantages over previous experiments. Foremostly, in the horse-race paradigm participants are not explicitly asked to evaluate a certain trait of the speaker (in this case, knowledgeability). Instead, take participants' bet distribution as an indirect measurement of their perceptions of knowledgeability. Indeed, previous studies have shown that individuals are sensitive to this manipulation and that disfluent information leads to smaller bets (e.g., Butterworth, 2019). Secondly, horse races provided a sensible context: It is a scenario where individuals can make decisions based on what they are told, but it is a rather obscure context where the content of speech itself may not be informative for most individuals (in that horses' descriptions are filled with jargon not comprehensible for those not familiar with the game).

Here, we present participants with a native and a non-native speaker, each describing two horses, with one description produced fluently and one description produced disfluently. If listeners are sensitive to both local and global causes of hesitations when making judgements about trust, then we expect that disfluent descriptions provided by a native speaker will result in less money bet, reflecting listeners' lower FOAK for the speaker. However, disfluent descriptions provided by a non-native speaker should not impact listeners' betting behaviour to the same degree if they consider the possibility of difficulties in production when assessing the speaker's knowledge.

To further control for the potential effect of (non)-nativeness on believability on its own, we will measure participants' language attitudes towards each speaker (see Dragojevic & Giles, 2016), the perceived fluency, accentedness, and comprehensibility of the native and the non-native speaker, as well as the perceived reliability of each speaker. Finally, we will measure participants' familiarity with and exposure to native and non-native-accented English on a daily basis, to account for the fact that exposure to non-native accents can reduce their negative effects on listeners' judgments (Boduch-Grabka & Lev-Ari, 2021).

2 Methods

All experimental stimuli, including validation data, can be found here.

2.1 Participants

We conducted a sensitivity power analysis via data simulation following DeBruine and Barr (2021) for our effect of interest, i.e., the interaction between fluency and speaker identity. We explored the required number of participants for a power of .8 in a 1000 simulations, varying the effect size of this interaction and the standard deviation of the residuals. In this analysis, we assumed a medium effect size for fluency, and no effect for speaker identity, following our hypothesis. This analysis showed that a sample size of 360 participants will ensure enough power from a medium effect size of the interaction ($\beta = 30$) onwards and a relatively large standard deviation of residuals. The simulation code can be found at here; note that the simulated data file is too large for GitHub, so the tile graph shows the power distribution as a function of number of participants (per group), effect size of the interaction, and standard deviation of the residuals.

Data collection will thus continue until we reach 360 participants who meet the following criteria: To have been born and raised in the United Kingdom, currently reside in the United Kingdom, be monolingual English speakers, have no auditory disorders, and have no previous experience or knowledge about horse races and betting on those, and to not have failed the experiment's attention checks. We controlled for linguistic background to control for potential confounding effects on accent perception (Grey, Schubel, McQueen, & Van Hell, 2019). Participants will be recruited from the online platform Prolific and will be reimbursed £1.5 in exchange for a 5-minute experiment and will be debriefed afterward.

2.2 Visual stimuli

We selected a set of eight images of racehorses obtained from the web. The selected images contained only one racehorse in the foreground, in motion, ridden by a jockey, and they all took up approximately the same amount of the image. We recruited ten participants on Prolific, who will not take part in the study, were asked to rate on a 10-point scale how likely each horse was to win a hypothetical race, and to rank the horses in the order they thought they would cross the finish line in exchange of £0.45.

A one-way repeated measures ANOVA showed that there were no differences in how likely each horse was thought to win a race (F(7) = 0.5, p = .83). An ordinal logistic regression showed that none of the horses were more likely to be ranked differently from the others (all $|t| < 2$).

2.3 Auditory stimuli

We selected four descriptions of actual race horses, originally used in Butterworth (2019), and retrieved from *The Racing Post* in October 2018. Each description consisted of three to four sentences describing the horse and its performance in previous races. We asked 30 British English speakers, who will not take part in the study, to rate Butterworth's (2019) original set to ensure that all descriptions were perceived as equally likely to describe a winning horse. Participants rated on a 10-point scale how likely they thought each horse was to win a race individually, and then ranked all the horses in exchange of £0.5. The results of this validation showed that some descriptions were more likely to be taken as reflecting winning horses.

Based on the results of this task, we further edited the descriptions and tested them in a new sample of 30 British English speakers. A one-way repeated measures ANOVA showed no differences in each horse's rated likelihood to win a race (F(3) = 0.76, p = .52), nor in the order in which they were ranked (all $|t| < 2$). The final set of descriptions for the experiment can be found in Table 1. Each description was paired with one of the four visual stimuli.

We recorded a British English native speaker and a non-native (L1: Italian) English speaker for our experiment. Both speakers were female. Passages were recorded one at a time. To elicit naturally disfluent recordings, both speakers were instructed to read the sentences silently and then were recorded as they tried to recall the passage from memory. To avoid differences between the descriptions provided by the speakers, they were allowed to look at the descriptions as they spoke if they could not remember the continuation. We edited these recordings using Audacity to ensure a similar amount of disfluencies between both speakers and at similar places, by cross-splicing different recordings. To create their fluent counterpart, filled and mid-utterance silent pauses were excised, and between-clauses silent pauses and elongations were reduced using the 'Tempo' function on Audacity. The final auditory experimental stimuli consisted, for each speaker, of two descriptions

Table 1: Breakdown of horse descriptions.

Horse	Description
Fire Walker	Fire Walker is looking strong thanks to his come-from-behind success in the Acomb Stakes. The impression given in both runs is that Fire Walker should handle the demands of the extra furlong and Charlie Hills is looking forward to the test. The trainer said “He’s done really well for a little break, his work’s been good and I couldn’t be more pleased with him”.
Silver Sky	Silver Sky, a runner-up of a seven-furlong maiden at Naas on his debut last month, the son of Invincible Spirit ran crack French colt Persian King to a neck in the Group 3 Autumn Stakes over today’s trip at Newmarket two weeks ago and his trainer believes he has done well since. O’Brien said “Silver Sky is a fine big colt and a talented one”.
Apocalypse	Apocalypse has put in a string of consistent performances, most recently finishing third to Norway in the Zetland. “He’s had a very solid year” said trainer Archie Watson. ”He ran a good race in the Zetland, beaten only a length and a quarter, and I think the field here is of a similar level so I’m more than happy for him to take his chance”.
Black Blade	Black Blade proved the market all wrong as the complexion of the 6.5-furlong novice race changed dramatically in the final two furlongs, with the Rebel Racing premier-owned newcomer under Tom Queally collaring long-time leaded Monsieur Noir. Spencer said: “He did it well. He’s a nice horse. We always thought had a bright future”.

of each horse (one disfluent, one fluent), resulting in sixteen recordings. Description, fluency, and speaker were counterbalanced in a Latin Square design resulting in 24 lists.

To ensure that the resulting descriptions were perceived to be natural (i.e., out edited ‘fluent’ audios were not clearly edited) and that they were matched in fluency (i.e., disfluent and ‘fluent’ versions were distinguishable), we validated them in a sample of 48 British English participants who will not take part in the study. Participants were allocated to one of the 24 experimental lists, to ensure that our validation procedure was similar to how participants will encounter stimuli in the actual experiment. Following Bosker, Quené, Sanders, and de Jong (2014) procedure, participants were asked to rate each audio’s fluency on a scale from 1 to 9 (1: not fluent at all, 9: very fluent). We instructed participants to rate fluency by considering silent and filled pauses, speed of speech and repairs, and to ignore speakers’ accents and the content of their speech. Participants additionally rated on a 9-point scale each recording’s naturalness (defined as how likely it was that the audio had been recorded in one go; 1: not unlikely at all; 9: very likely), and accentedness (while ignoring the perceived speaker’s proficiency of the language; 1: not accented at all, 9: very accented). We additionally asked participants to guess the speakers’ country of origin. At the end of the task,

Table 2: Mean (standard deviation) rating of native and non-native speakers' fluent and disfluent recordings for fluency, naturalness, and accentedness, on a 9-point scale where lower values indicate less fluent, less natural, and less accent respectively.

	Fluency	Naturalness	Accentedness
Native speaker			
Fluent	7.62 (1.59)	5.44 (2.62)	4.38 (2.22)
Disfluent	6.44 (2.36)	5.73 (2.57)	4 (1.97)
Non-native speaker			
Fluent	6.25 (1.73)	5.94 (1.83)	7.02 (1.59)
Disfluent	5.21 (1.75)	5.83 (1.99)	7.17 (1.51)

participants were further asked how often they interact with native and non-native English speakers (on a 9-point scale, 1: never, 9: always) and were allowed to report if they noticed anything odd in the auditory stimuli.

Table 2 shows the mean (and standard deviations) of participants' ratings in fluency, naturalness, and accentedness. A linear mixed model for participants' fluency ratings as predicted by fluency, speaker's linguistic background, and their interaction, with random intercepts by-participant and by-horse description showed that fluency ratings differed significantly for our fluent and disfluent conditions (β = 1.19, SE = 0.34, t = 3.52). Although the non-native speaker was perceived as more disfluent than the native speaker (β = -1.23, SE = 0.34, t = -3.64) (in line with previous findings, e.g., Pinget, Bosker, Quené, & De Jong, 2014; Bosker, Quené, Sanders, & de Jong, 2014), the interaction between the two variables was not significant (β = -0.15, SE = 0.48, t = -0.31). An identical model for naturalness ratings showed no significant differences by fluency (β = -0.29, SE = 0.41, t = -0.71), speaker's linguistic background (β = 0.10, SE = 0.41, t = 0.26) or their interaction (β = 0.40, SE = 0.58, t = 0.69).

2.4 Procedure

Stimuli are presented using JsPsych (de Leeuw, Gilbert, & Luchterhandt, 2023), hosted on MindProbe (via JATOS, Lange, Kühn, & Filevich, 2015). The task begins with a cover story that introduces two horse racing tipsters, who will provide information about the four most popular horses who will compete in an upcoming race at Musselburgh Racecourse (Edinburgh). This cover story explains that the two tipsters are well-known experts in the field, and explains that one of the speakers is a non-native English speaker (without specifying the nationality of either speaker), introducing the element of the speakers' linguistic backgrounds as well as the factor of competence.

At the beginning of the experiment, participants are shown the four pictures of the horses that will take part in the race, alongside their names. Participants are instructed to distribute one-hundred pounds across the four horses based on the likelihood they thought each horse had of winning: They can split the bets as they wish, and they do not have to spend all the money. In this screen, participants are informed that they will listen to either speaker describe each horse only once and that they will place their bet on each horse after each description. Each participant is randomly assigned to one out of 24 groups, so that they will listen to each speaker twice, one in each fluency level. Horse presentation is randomised. In each trial, participants listen to one speaker provide a description of the horse's performance. Once the recording stops, participants are asked to place a bet by typing a number on a web form. Participants can only move to the next horse's descriptions once they have placed a bet. Participants are allowed to modify their previous bets every time they hear a new description. If participants bet more than the allotted maximum, they are asked to re-distribute their bets until the total is below one-hundred pounds.

After the experiment, participants will complete a questionnaire following that in Foucart et al. (2021) to measure their language attitudes towards the native and the non-native speaker. Per speaker, participants have to answer six questions measuring affect (three questions for negative affect and three questions for positive affect), five questions measuring solidarity, five questions measuring status, and one question for comprehensibility, accentedness, fluency, and trustworthiness – all on a 9-point scale. Set presentation (native/non-native) will be randomised across participants, with the order of presentation of dimensions being randomised. We will also ask participants to guess the country of origin of our native and non-native speakers, and from what speaker they would like to learn about horse races in the future. Additional questions include ratings on a 9-point scale of how natural the audio sounded (1: unnatural, edited; 9: natural, unedited) and our participants' exposure to native and non-native accented English (1: never; 9: always). Likewise, we measure participants' previous experience with betting and their perceived knowledge of horse races (two questions: Whether they had bet on horseraces in the past, and to rate on a 5-point scale - from 'Strongly disagree' to 'Strongly agree' - how much they identify with the statement 'I am an expert in horseraces'). We will include an open-ended question for participants to report what guided their decision-making as well as the experiment's aim. If participants report having knowledge of horseraces (i.e., more than 'Neither agree nor disagree' with the statement 'I am an expert in horseraces'), rate the naturalness of the audio lower than four, or directly guess the experiment's aim (they report the auditory manipulation) they will be discarded from the analysis.

3 Analysis

All analyses will be carried out in R (version 4.3.1., R Core Team, 2023), with the packages *tidyverse* (version 2.0.0, Wickham et al., 2019) for data wrangling, *ggplot2* (version 3.4.2, Wickham, 2016) for data visualization, and *lme4* (version 1.1.33, Bates, Mächler, Bolker, & Walker, 2015) for analyses. Scripts for data wrangling, visualization, and analysis can be found at OSF repository and GitHub repository.

3.1 Betting behaviour

Participants' betting behaviour will be modelled using a linear mixed model. The model will include fixed effects of fluency (sum-coded, fluent coded as -0.5, disfluent as +0.5), speaker's linguistic background (sum-coded, native coded as -0.5, non-native coded as +0.5), and their interactions. We will model the maximal model (Barr, Levy, Scheepers, & Tily, 2013). This model includes random intercepts by-participant and by-item, with random slopes for fluency and speaker's linguistic background by-participant, and for fluency by-item. Results will be deemed significant at $|t| > 2$ (Baayen, 2008). If our model fails to converge, we will simplify it by sequentially dropping the factors where (1) there is a perfect correlation in the random effects and (2) explain the least variance.

3.2 Language attitudes

As part of an exploratory analysis, we will obtain participants' scores on Affect, Status, and Solidarity by calculating their average per participant. In the case of the dimension Affect, we will reverse score the items measuring negative affect. We will first explore whether there were differences between the native and non-native speakers in these three dimensions, as well as on Comprehensibility, Accentedness, and Trustworthiness via paired t-tests using Bonferroni correction for p values. Whenever there is a difference in scores between the native and the non-native speaker, we will explore whether participants' betting behaviour could be further accounted for by this difference. We will do so by including the variable in a new model, identical to the one described in section 3.1.

4 Appendix: Experiment instructions

4.1 Task instructions

‘Scottish Flat Season Finale’ race

We have received information from two leading horse racing tipsters about the upcoming S̈cottish Flat Season Finaler̈ace next Monday, 16th October at Musselburgh Racecourse (Edinburgh).

Our two speakers are highly respected tipsters amongst the horse racing elites (note that one of our speakers is a non-native English speaker). Both have achieved an outstanding level of performance that marks them as the very best in the game. They currently stand undefeated, having had an unbeatable strike of consecutive wins since 2021.

We are truly honoured to have received information on the 4 most popular horses running at Musselburgh in October, and are looking forward to sharing this information with you.

Betting slip screen

These are the four horses we have received information about. You can see the name and a picture of each.

During the study, you will hear four descriptions, one about each horse. You will only hear each passage once. You have **one-hundred** virtual pounds to split between the four horses. You must place a bet on each horse based on the likelihood you think each horse has of winning. You may split the money however you like, and you do not have to spend it all. You can place your bets as you go, and you can edit previous bets.

There will be further instructions on the screen, which should make your task clear.

Individual betting screen

This is your [first/second/third/forth] bet. Enter a quantity, out of a hundred, in the box under the horse picture. Please, use digits instead of text. You can submit your response once the audio is over, by clicking on 'Continue' below. You will then listen to the next horse description. You have [X] pounds left.

Exceeded maximum allowed screen

You bet more than your allocated 100 pounds (your current bet is [X]). Please, re-distribute your bets. You allocated [horse name and last amount allocated].

5 Appendix: Questionnaire

Language attitudes questions (from Foucart et al., 2021). Note that 'this speaker' is replaced by the speaker's linguistic background in the experiment.

Affect

- How annoyed did you feel when you listened to this speaker?
- How irritated did you feel when you listened to this speaker?
- How frustrated did you feel when you listened to this speaker?
- How interested did you feel when you listened to this speaker?
- How happy did you feel when you listened to this speaker?
- How enthusiastic did you feel when you listened to this speaker?

Status

- How intelligent did you find this speaker?
- How educated did you find this speaker?
- How smart did you find this speaker?
- How competent did you find this speaker?
- How successful did you find this speaker?

Solidarity

- How friendly did you find this speaker?
- How nice did you find this speaker?
- How pleasant did you find this speaker?
- How honest did you find this speaker?
- How sociable did you find this speaker?

Accent strength

- How strong was this speaker's accent?

Comprehension

- How easy was it to understand this speaker?

Fluency

- How fluent was this speaker?

Additional item:

Trustworthiness

- How trustworthy was this speaker?

6 References

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r.* Cambridge University Press. doi: 10.1017/CBO9780511801686

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beattie, G. W., & Butterworth, B. (1979, 7). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*(3), 201–211. Retrieved from `https://doi.org/10.1177%2F002383097902200301` doi: 10.1177/002383097902200301

Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive science*, *45*(11), e13064. doi: 10.1111/cogs.13064

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116. doi: 10.1016/j.jml.2014.05.004

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). The perception of fluency in native and nonnative speech. *Lang. Learn.*, *64*(3), 579–614. doi: 10.1111/lang.12067

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383–398. doi: 10.1006/jmla.1995.1017

Butterworth, H. (2019). *How much do they, um, know?: Disfluencies in speech as cues to speakers' knowledge.* (Unpublished MA thesis)

Caballero, J. A., & Pell, M. D. (2020). Implicit effects of speaker accents and vocally-expressed confidence on decisions to trust. *Decision*, *7*(4), 314. doi: 10.1037/dec0000140

DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920965119. doi: 10.1177/251524592096511

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jspsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351. doi: 10.21105/joss.05351.

Dragojevic, M., Giles, H., Beck, A.-C., & Tatum, N. T. (2017, 5). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385–405. doi: 10.1080/03637751.2017.1322213

Foucart, A., Costa, A., Morís-Fernández, L., & Hartsuiker, R. J. (2020). Foreignness or processing fluency? On understanding the negative bias toward foreign-accented speakers. *Language Learning*, *70*(4), 974–1016. doi: 10.1111/lang.12413

Foucart, A., & Hartsuiker, R. J. (2021, 7). Are foreign-accented speakers that 'incredible'? The impact of the speaker's indexical properties on sentence processing. *Neuropsychologia*, *158*, 107902. doi: 10.1016/j.neuropsychologia.2021.107902

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709–738. doi: 10.1006/jmla.1995.1032

Gregersen, T. S. (2005, 10). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, *38*(3), 388–400. Retrieved from `https://doi.org/10.1111%2Fj.1944-9720.2005.tb02225.x` doi: 10.1111/j.1944-9720.2005.tb02225.x

Grey, S., Schubel, L. C., McQueen, J. M., & Van Hell, J. G. (2019). Processing foreign-accented speech in a second language: Evidence from ERPs during sentence comprehension in bilinguals. *Bilingualism: Language and Cognition*, *22*(5), 912–929. doi: 10.1017/S1366728918000937

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012, 4). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. doi: 10.1162/jocn_a_00103

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2019, 8). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1), 55–79. doi: 10.1177/1747021819865833

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126. doi: 10.1016/j.specom.2017.01.011

Lange, K., Kühn, S., & Filevich, E. (2015). " just another tool for online studies"(jatos): An easy solution for setup and management of web servers supporting online studies. *PloS one*, *10*(6), e0130834. doi: 10.1371/journal.pone.0130834

Lev-Ari, S., & Keysar, B. (2010, 11). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. doi: 10.1016/j.jesp.2010.05.025

Mai, R., & Hoffmann, S. (2013, 9). Accents in business communication: An integrative model and propositions for future research. *Journal of Consumer Psychology*, *24* (1), 137–158. doi: 10.1016/j.jcps.2013.09.004

Matzinger, T., Pleyer, M., & Żywiczyński, P. (2023). Pause length and differences in cognitive state attribution in native and non-native speakers. *Languages*, *8* (1), 26. doi: 10.3390/languages8010026

Pinget, A. F., Bosker, H. R., Quené, H., & De Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31* (3), 349–365. doi: 10.1177/026553221452

Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory control and L2 proficiency modulate bilingual language production: Evidence from spontaneous monologue and dialogue speech. *Frontiers in Psychology*, *3*. Retrieved from `https://doi.org/10.3389%2Ffpsyg.2012.00057` doi: 10.3389/fpsyg.2012.00057

R Core Team. (2023). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60* (3), 362. doi: 10.1037/0022-3514.60.3.362

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & language*, *25* (4), 359–393. doi: 10.1111/j.1468-0017.2010.01394.x

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4* (43), 1686. doi: 10.21105/joss.01686