Social Robotics Meets Sociolinguistics: Investigating Accent Bias and Social Context in HRI

Mary Ellen Foster
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
MaryEllen.Foster@glasgow.ac.uk

Jane Stuart-Smith
Glasgow University Laboratory of Phonetics
University of Glasgow
Glasgow, United Kingdom
Jane.Stuart-Smith@glasgow.ac.uk

ABSTRACT

Deploying a social robot in the real world means that it must interact with speakers from diverse backgrounds, who in turn are likely to show substantial accent and dialect variation. Linguistic variation in social context has been well studied in human-human interaction; however, the influence of these factors on human interactions with digital agents, especially embodied agents such as robots, has received less attention. Here we present an ongoing project where the goal is to develop a social robot that is suitable for deployment in ethnically-diverse areas with distinctive regional accents. To help in developing this robot, we carried out an online survey of Scottish adults to understand their expectations for conversational interaction with a robot. The results confirm that social factors constraining accent and dialect are likely to be significant issues for human-robot interaction in this context, and so must be taken into account in the design of the system at all levels.

CCS CONCEPTS

• **Social and professional topics** → **Cultural characteristics**; • **Human-centered computing** → **Natural language interfaces**; **User studies**.

KEYWORDS

digital agents, sociolinguistics, social robotics

ACM Reference Format:
Mary Ellen Foster and Jane Stuart-Smith. 2023. Social Robotics Meets Sociolinguistics: Investigating Accent Bias and Social Context in HRI. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3568294.3580063

1 INTRODUCTION

When a social robot is deployed in the real world, it has to interact with speakers from diverse backgrounds, who in turn are likely to show substantial accent and dialect variation. It is already well known that speech recognition in (non-embodied) digital agents performs more poorly for speakers of non-standard dialects [11]. However, while public-space interaction is an increasingly well-studied problem in Human-Robot Interaction, and there have been a few recent results on sociophonetic interaction with digital agents [31], less is known about the specific role of sociolinguistic variation in the context of HRI.

HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden
2023. ACM ISBN 978-1-4503-9970-8/23/03…$15.00
https://doi.org/10.1145/3568294.3580063

Our goal is to develop a robot that can be deployed in real-world contexts, and so in socially and ethnically-diverse areas with distinctive regional accents, which characterises most contemporary urban settings. We aim to design, and then refine, a conversational system for the robot to function most effectively in this sociolinguistic context, meeting users' expectations and supporting fluid natural-language conversations with a wide range of interaction partners. As part of the robot design process, we conducted an online study of 151 Scottish participants to measure their *a priori* expectations about spoken interactions with the robot. The survey covers various aspects of user expectations, including factors such as whether the participants expect the robot to understand them easily and whether they anticipate changing their speech when talking to the robot. The survey also gathered a range of demographic factors from the participants, which we use to inform the impact of social factors on users' expectations.

2 BACKGROUND

2.1 How do humans talk to each other?

To appreciate how humans might respond to artificial agents, we need first to consider the wealth of information from sociolinguistic research into what happens when humans talk to each other, since it is likely that humans extend and adapt behaviours evolved for interpersonal interaction to that with non-human agents [19]. Variationist sociolinguistics has empirically demonstrated systematic relationships between linguistic variation and group-level social factors, including age, gender, region, social class and ethnicity [12, 28]. Related to this, many languages show systematic distinctions between supralocal 'standard' varieties, associated with prestige, education, professional roles, and national media and broadcasting, and 'non-standard' dialects, associated with local traditional regional working-class communities, and/or migrant ethnic communities, and often represented in entertainment media [18].

Intraspeaker linguistic variation, analysed under 'speech style', is also systematic, responsive to a number of factors and explained by several complementary theories [9]: speakers may monitor their speech especially when in more formal situations, shifting to more standard linguistic variation; they may accommodate their speech patterns to those of their interlocutor, or when this is phonologically or socially incongruent, may diverge; they may design their speech for their audiences, both present and imagined; and they may construct locally-meaningful sociolinguistic identities, which enable speakers to affiliate or differentiate themselves from others. Speakers and listeners are also highly sensitive to social information

present in the speech signal, often displaying accent bias ('accentism') which typically favours standard over non-standard accents [13], in turn influencing cognitive and linguistic processing [21, 32].

2.2 How do humans respond to artificial agents?

Users are known to form initial impressions of an agent based on how it talks, even quite a subtly; for example, Torre et al. [30] compared user behaviour when interacting with virtual partners that used either Southern Standard British English or a regional variety, and found that users' trust in the agent appeared to be affected by the initial impression of the agent formed from its dialect, echoing the results of how humans evaluate standard and regional UK accents [10]. In the context of physically embodied robot agents, while a small number of papers have directly investigated the role of accent [15, 29], the influence of this factor remains under-explored.

Understanding the interaction between humans and digital agents rests on many variables, including: the design of the agent's interaction capabilities (e.g. script, voice/accent), the capacity for the automatic speech recogniser (ASR) to recognize voices/accents, the responses of humans to computers in general [19], and digital agents in particular [6], ranging from listener preferences for certain voice types or personae [7], to speaker responses to, and alignment with, digital agents' speech [4, 23], and listeners extending existing accent bias to digital voices [16]. The perceived gender of the agent can also affect users' interactions with the agent, and an increasing body of work is examining how these gendered expectations influence users' interactions with virtual agents [1].

Despite all of these factors, modern digital agents usually interact in standard language, using voices with standard accents (Southern Standard British English, General American) and are equipped with ASR systems whose performance can be trained to individual interlocutors, but tend to perform better with standard dialects. This is partly for accessibility reasons, with the goal of supporting the widest possible number of users, and partly as a result of biases within the speech technology industry and NLP more generally [5, 11]. However, note that this is starting to change: Apple's Siri has recently added a range of new default American voices, and also no longer defaults to a gendered voice [20], while the BBC's 'Beeb' voice assistant has been given a Northern accent [3].

In contrast with the increasing work on how users form impressions based on a virtual agent's speech, relatively little is known about how humans styleshift towards digital agents, though there is evidence for fine-grained adjustments in speech rate to an avatar, and even more so if the speaker identifies with the agent, even whilst knowing it is an avatar [7, 24]. There is also evidence that speakers will adapt their speech towards a virtual agent in the same way that they would adapt to a non-native speaker [14].

3 SOCIOLINGUISTIC EXPECTATIONS FOR HRI

The overall goal of this project is to develop an English-speaking robot that can be deployed successfully in public spaces in an urban context known to show socially and ethnically diverse, non-standard accents; for the initial study, we specifically consider the Scottish city of Glasgow. For this robot, as for all public robots, it is crucially important that users are able to interact with it with as few barriers as possible. For this reason, and in the absence of research on the sociolinguistics of HRI, we began the design process by conducting an online survey of Scottish adults to understand their expectations about human-robot interaction, and also to determine how those expectations are influenced by social factors.

Scotland, and in particular, Scotland's largest city, Glasgow, represents an excellent case study for a potentially challenging real-world sociolinguistic situation for HRI. Glasgow is Scotland's most socially and ethnically diverse city, with much of the population speaking a dialect of Scottish English, also known as 'Scots', which historically continues a form of Northern Anglian, and which is different in terms of accent, grammar and vocabulary, from Standard Scottish English, which continues Southern English adopted into Scotland since the 17th century [8]. Glasgow Scots holds substantial covert prestige and acts as a strong marker of solidarity amongst working class speakers [17, 27], though many middle-class speakers can also switch, or drift, between Scots and Standard Scottish English depending on interlocutor and context, with more formal contexts provoking the shifts to the overtly prestigious standard [25]. Glasgow is also home to a number of minority ethnic communities, notably the Glasgow-Asian ('Glaswasian') and Scottish-Polish heritage communities, both of whose accents have been researched [22, 26], as well as refugee and asylum seeker communities.

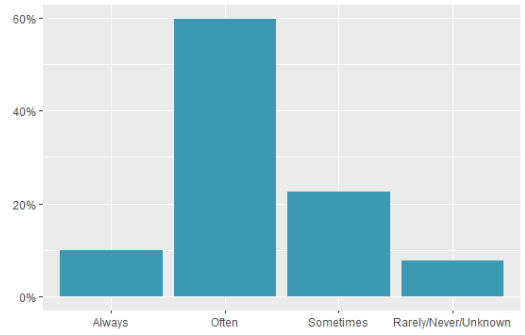
The Glasgow dialect is notoriously difficult for speakers of other English dialects to understand [2]. Anecdotally, Scottish speakers complain that digital agents cannot understand their accents, as captured by the well-known BBC Scotland 'Voice Activated Elevator' comedy sketch (https://www.bbc.co.uk/programmes/p00hbfjw) and also in comments such as this from our recent survey:

- '... When using speech recognition like work dictate and Alexa and Siri they don't really get the gist of what I am saying .' (Female, 30-34, Middle class)

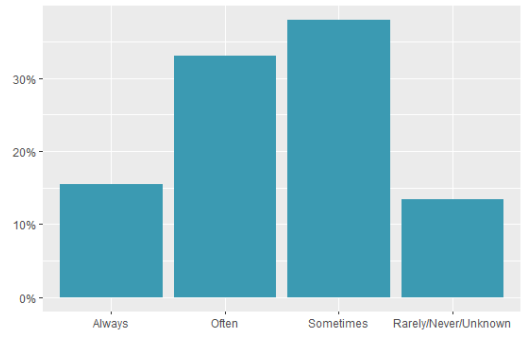
3.1 The survey

To determine user expectations about the robot, an anonymous online survey was conducted. Participants were recruited through social media, with a particular outreach to Scottish participants. The survey asked a number of questions about users' previous experience interacting with artificial agents and robots, their expectations for interacting with robots in the future, and what factors they anticipated might affect the quality of those interactions; for many of the questions, a free-form box was provided to gather additional details about the responses. At the end of the survey, users were asked to answer a range of core demographic questions, and were then invited to give any additional free-text comments on the general topic of human-robot interaction in public spaces. The full list of survey questions is available online at <https://www.dcs.gla.ac.uk/~mefoster/survey_questions_hri2023_lbr.pdf>.

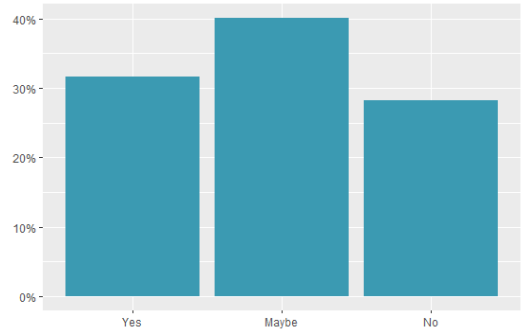
The survey had 151 participants. The majority were Scottish residents, largely from the Glasgow area. Around two-thirds of the participants identified as female, with a broad age distribution. Approximately half of the participants (75 of 151) self-identified as working class, while the others either identified as middle class or indicated that they did not believe social class matters.



(a) 'I will be able to understand the robot.'



(b) 'I will have to change the way I speak for the robot to understand me.'



(c) 'I believe having a conversation with the robot will be easy.'

Figure 1: Summary of relevant questionnaire responses

3.2 Results

3.2.1 Quantitative analysis. For this analysis, we concentrated on users' responses to the following items from the survey, as they directly address the overall research goal of assessing users' expectations of conversations with a robot.

- *I will be able to understand the robot.*
- *I will have to change the way I speak for the robot to understand me.*
- *I believe having a conversation with the robot will be easy.*

The overall pattern of responses to these three survey items are shown in Figure 1. In general, the study participants did not anticipate any difficulty understanding the robot (Figure 1a), but largely felt that they would need to change they way they spoke for the robot to understand them (Figure 1b). Responses on the final item showed a clearer pattern, in that 70% of respondents expressed doubt ('Maybe'/'No') about whether having a conversation with the robot would be easy (Figure 1c).

We then examined the influence on these results of three main social factors: Age, Social class, and Gender identity. The graphs in Figures 2–4 present the results broken down by these three factors. For each question, we present the descriptive patterns in the results, and then identify any significant effects. Note that not all of these patterns were found to be statistically signficant, likely due to the complexity of interactions across the social factors as shown on the graphs.

Understanding the robot The detailed responses for this question are shown in Figure 2. In general, younger participants were more confident about understanding the robot than older participants, across all gender and social class combinations, while it also appears that both middle-class respondents and male respondents gave generally more positive answers on this question. The influence of age was found to be significant on a χ^2 test: $\chi^2 = 15.071$, $df = 6$, $p < 0.05$)

Changing speech The detailed responses for this question are shown in Figure 3. The main factor of note on this question was that, for male participants only, the participants' age had a very significant effect on their responses to this question ($\chi^2 = 19.596$, $df = 6$, $p < 0.005$), with young men the least likely to expect to change their accent. There are also two non-significant trends: female participants generally anticipated having to change their speech style more, and social class affects reponses in interaction with gender for young participants, but in opposite directions.

Ease of interaction The detailed responses for this question are shown in Figure 4. Young men appear to have been more positive about the interactions with the robot than young women, while for the older age groups, the responses from female participants appear to have been more confident. As on other questions, the working-class respondents, especially young females and young and middle-aged males, were also less positive about the ease of interacting with the robot.

3.2.2 Qualitative analysis. We also carried out a qualitative analysis of the free-form response data, which reveals several recurrent themes echoing the general patterns from the quantitative analysis above. Concern about the robot's ability to understand the participants' accent is a prevalent theme, especially in respondents identifying as working-class, indicating that sociolinguistic factors, especially relating to non-standard dialects, are likely to affect Human Robot Interaction. In fact, 30 of the participants (mainly self-identified working class) explicitly mentioned accents when elaborating on a 'No' or 'Maybe' response to the question about ease of interaction; some representative responses include:

- 'Would the robot understand my accent? If not, it would mean repeating myself over and over.' (Male, 50-54, Working class)
- 'Difficulty understanding accent may lead to having to repeat myself for the robot to understand me, or it may never end up understand me.' (Female, 20-25, Working class)
- 'Its ability to understand my accent for certain words.' (Female, 35-39, Working class)

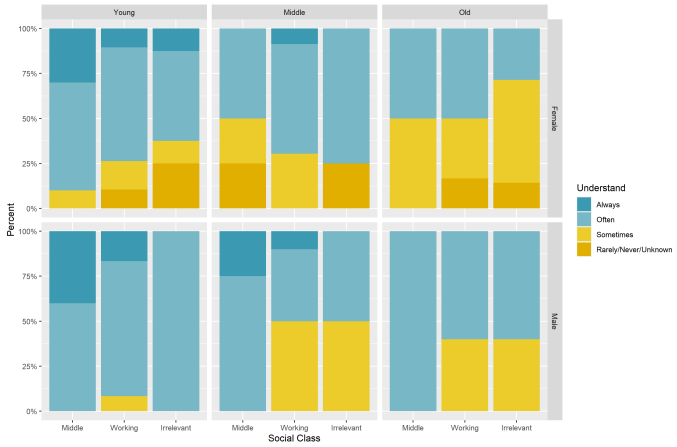


Figure 2: Detailed responses to 'I will be able to understand the robot'

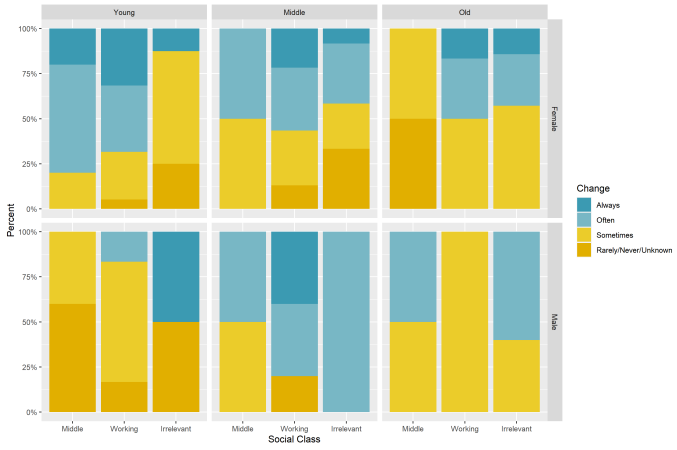


Figure 3: Detailed responses to 'I will have to change the way I speak for the robot to understand me.'

In response to a later question asking about experiences with conversational agents, some participants raised similar issues. For example, when referring to a voice-enabled navigation system, one participant said:

- 'It didn't understand my accent and after a few attempts it ignored me and said 'Closing Down' in a snotty English accent, and I ended up swearing at it.' (Female, 55-59, Class not disclosed)

Note however that many participants also reported positive past experiences with conversational systems and robots, with many specifically referring to the 'fun' aspects of these interactions.

4 SUMMARY AND FUTURE DIRECTIONS

The underlying motivation for this project is to increase the understanding of how insights from sociolinguistics regarding human-human conversation transfer to the human-robot situation, particularly when it comes to social factors relating to accent and

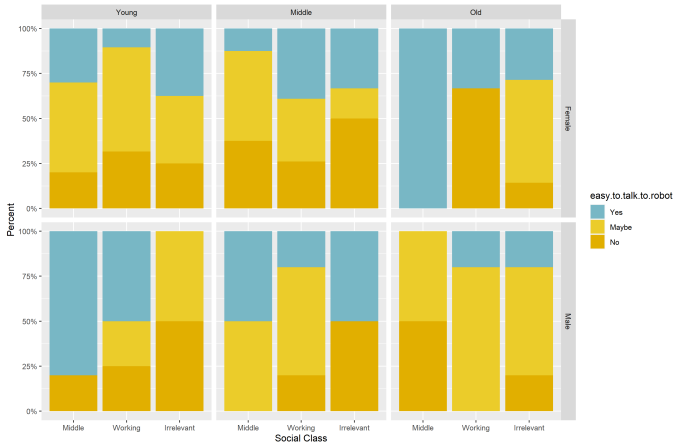


Figure 4: Detailed responses to 'I believe having a conversation with the robot will be easy?'

dialect for both human and robot. To study this, we are developing a public-space social robot with the goal of deploying it regularly in a range of sociolinguistic contexts. To gather user expectations of this sort of robot, we carried out an online survey to gather user expectations about HRI from Scottish adults: Scotland represents a challenging real-world sociolinguistic situation for this HRI, with a range of socially and ethnically diverse, non-standard dialects.

Our results reveal a sociolinguistic asymmetry for HRI: potential robot users expect to be able to understand the robot (likely assuming that it will speak standard English), but they also assume that the robot might have difficulty understanding them, and that they might need to change the way they speak to have a successful interaction. Moreover, participant responses are modulated by interacting social factors of age, gender and social class. Misunderstanding of accent and dialect was also mentioned frequently as a potential issue in the free-text responses from this largely Scottish sample. These findings indicate, along with those for digital agents [16, 31], that interaction with embodied agents is also subject to the kinds of sociolinguistic constraints which govern human-human communication [c.f. 19], and hence opens up our understanding of interaction as a multidimensional continuum which extends and encompasses a range of interlocutors, human and agent.

The immediate next step in this project is to develop an initial robot system and to carry out sociolinguistic observation of user interaction and reactions to it in public spaces. This survey has given some suggestions as to the user expectations. Observing actual interactions will provide input on more specific design decisions: for example, will users style-shift towards the robot? Is this behaviour affected by whether the robot is given a regionally accented voice? And, how are these (and other) behaviours influenced by social factors? Follow-up interviews with participants to understand their expectations and concerns more deeply will also help to develop a robot that meets the expectations and requirements for public-space interaction in real-world contexts.

ACKNOWLEDGMENTS

This work was partially supported by the UKRI centre for Doctoral Training in Socially Intelligent Artificial Agents, Grant Number EP/S02266X/1. Thanks to Rhiannon Fyfe for the design and implementation of the online survey.

REFERENCES

[1] Gavin Abercrombie, Amanda Cercas Curry, Mugdha Pandya, and Verena Rieser. 2021. Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants. In *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*. Association for Computational Linguistics, Online, 24–33. https://doi.org/10.18653/v1/2021.gebnlp-1.4

[2] Patti Adank, Bronwen G Evans, Jane Stuart-Smith, and Sophie K Scott. 2009. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of experimental psychology. Human perception and performance* 35, 2 (2009), 520–529. https://doi.org/10.1037/a0013552

[3] BBC News. 2020. OK Beeb: BBC voice assistant will learn regional accents. (2020). https://www.bbc.co.uk/news/technology-52891155

[4] Linda Bell. 2003. Linguistic adaptations in spoken human-computer dialogues. *Kungliga Tekniska Högskolan, Stockholm* (2003).

[5] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. (July 2020), 5454–5476. https://doi.org/10.18653/V1/2020.ACL-MAIN.485 arXiv:2005.14050

[6] Michelle Cohn, Kai Hui Liang, Melina Sarian, Georgia Zellou, and Zhou Yu. 2021. Speech Rate Adjustments in Conversations with an Amazon Alexa Socialbot. *Frontiers in Communication* 6 (2021). https://doi.org/10.3389/FCOMM.2021.671429/FULL

[7] Michelle Cohn, Kristin Predeck, Melina Sarian, and Georgia Zellou. 2021. Prosodic alignment toward emotionally expressive speech: Comparing human and Alexa model talkers. *Speech Communication* 135 (Dec. 2021), 66–75. https://doi.org/10.1016/j.specom.2021.10.003

[8] John B Corbett, J Derrick McClure, and Jane Stuart-Smith. 2003. A brief history of Scots. In *The Edinburgh Companion to Scots*, John B. Corbett, J. Derrick McClure, and Jane Stuart-Smith (Eds.). Edinburgh University Press, 1–16.

[9] Nikolas Coupland. 2007. *Style: Language Variation and Identity*. CUP, Cambridge.

[10] Nikolas Coupland and Hywel Bishop. 2007. Ideologised values for British accents 1. *Journal of sociolinguistics* 11, 1 (2007), 74–93.

[11] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences of the United States of America* 117, 14 (April 2020), 7684–7689. https://doi.org/10.1073/PNAS.1915768117

[12] William Labov. 1972. *Sociolinguistic Patterns*. Blackwell, Oxford.

[13] Rosina Lippi-Green. 2012. *English with an accent: Language, ideology, and discrimination in the United States*. Routledge.

[14] Birgit Lugrin, Benjamin Eckstein, Kirsten Bergmann, and Corinna Heindl. 2018. Adapted Foreigner-Directed Communication towards Virtual Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (Sydney, NSW, Australia) *(IVA '18)*. Association for Computing Machinery, New York, NY, USA, 59–64. https://doi.org/10.1145/3267851.3267859

[15] Conor McGinn and Ilaria Torre. 2019. Can you Tell the Robot by the Voice? An Exploratory Study on the Role of Voice in the Perception of Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. https://doi.org/10.1109/hri.2019.8673305

[16] Zion Mengesha, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman. 2021. "I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans. *Frontiers in Artificial Intelligence* 4 (Nov. 2021), 169. https://doi.org/10.3389/FRAI.2021.725911/BIBTEX

[17] Janet Menzies. 1991. An Investigation of Attitudes to Scots. http://www.arts.gla.ac.uk/STELLA/STARN/lang/MENZIES/menzie1.htm

[18] James Milroy and Lesley Milroy. 2012. *Authority in language: Investigating language prescription and standardisation* (second ed.). Routledge, London.

[19] Clifford Nass, Janathan Steuer, and Ellen R. Tauber. 1994. Computer are social actors. *Conference on Human Factors in Computing Systems - Proceedings* (1994), 72–78. https://doi.org/10.1145/259963.260288

[20] Matthew Panzanario. 2021. Apple adds two brand new Siri voices and will no longer default to a female or male voice in iOS. *TechCrunch* (2021). https://techcrunch.com/2021/03/31/apple-adds-two-siri-voices/

[21] Donald L. Rubin. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33, 4 (Aug. 1992), 511–531. https://doi.org/10.1007/BF00973770

[22] Sadie Durkacz Ryan. 2021. "I just sound Sco[ʔ]ish now": The acquisition of word-medial glottal replacement by Polish adolescents in Glasgow. *English World-Wide* 42, 2 (April 2021), 145–174. https://doi.org/10.1075/EWW.00066.DUR

[23] Cathryn Snyder, Michelle Cohn, and Georgia Zellou. 2019. Individual variation in cognitive processing style predicts differences in phonetic imitation of device and human voices. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Vol. 2019-Septe. International Speech Communication Association, 116–120. https://doi.org/10.21437/INTERSPEECH.2019-2669

[24] K. Staum Casasanto, L Jasmin and D. Casasanto. 2010. Virtual accommodating: Speech rate accommodation to a virtual interlocutor. In *Proc. 32nd Ann Conf Cog Sci Soc., Austin*, S. Ohlsson and R. Catrambone (Eds.). 127–32.

[25] Jane Stuart-Smith. 2003. The phonology of Modern Urban Scots. In *The Edinburgh Companion to Scots*, John Corbett, Derek J McClure, and Jane Stuart-Smith (Eds.). Edinburgh University Press, 110–137.

[26] Jane Stuart-Smith, Claire Timmins, and Farhana Alam. 2011. Hybridity and Ethnic Accents: A Sociphonetic Analysis of 'Glaswasian'. In *Language Variation - European Perspectives III: Selected Papers from the 5th International Conference on Language Variation and Change in Europe (ICLaVE 5), Copenhagen, June 2009*, Frans Gregersen, Jeffrey K Parrott, and Pia Quist (Eds.). John Benjamins, 43–58.

[27] Jane Stuart-Smith, Claire Timmins, and Fiona Tweedie. 2007. Talkin' 'Jockney'? Variation and change in Glaswegian accent. *Journal of Sociolinguistics* 11, 2 (2007), 221–260. https://doi.org/10.1111/j.1467-9841.2007.00319.x

[28] Sali Tagliamonte. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Wiley-Blackwell, Oxford.

[29] Rie Tamagawa, Catherine I. Watson, I. Han Kuo, Bruce A. MacDonald, and Elizabeth Broadbent. 2011. The Effects of Synthesized Voice Accents on User Perceptions of Robots. *International Journal of Social Robotics* 3, 3 (June 2011), 253–262. https://doi.org/10.1007/s12369-011-0100-4

[30] Ilaria Torre, Jeremy Goslin, Laurence White, and Debora Zanatto. 2018. Trust in Artificial Voices: A "Congruency Effect" of First Impressions and Behavioural Experience. In *Proceedings of the Technology, Mind, and Society* (Washington, DC, USA) *(TechMindSociety '18)*. Association for Computing Machinery, New York, NY, USA, Article 40, 6 pages. https://doi.org/10.1145/3183654.3183691

[31] Georgia Zellou, Michelle Cohn, and Bruno Ferenc Segedin. 2021. Age- and Gender-Related Differences in Speech Alignment Toward Humans and Voice-AI. *Frontiers in Communication* 5 (2021). https://doi.org/10.3389/fcomm.2020.600361

[32] Yi Zheng and Arthur G. Samuel. 2017. Does seeing an Asian face make speech sound more accented? *Attention, Perception, and Psychophysics* 79, 6 (Aug. 2017), 1841–1859. https://doi.org/10.3758/S13414-017-1329-2/TABLES/5