Everyone has an accent

Nina Markl, Catherine Lai

University of Edinburgh, Scotland

`nina.markl@ed.ac.uk, c.lai@ed.ac.uk`

Abstract

In this paper, we consider how the notion "accent" in particular in the context of "accented speech" has been discussed in Interspeech publications between 2004 and 2022. We contrast the way speech technology research published in the conference has conceptualised these terms with their usage in linguistics. The point of this comparison is to: highlight significant inter-disciplinary differences in the way apparently core terms are used, discuss disadvantages of using inexact language in research, and encourage researchers to be more mindful about the use of particular short-hands.

Index Terms: accent, variation, accented speech

1. Introduction

As speech technologies become increasingly ubiquitous, especially for some "high-resource" (standard) languages and their speakers, "accent variation" has become a popular research topic at speech technology conferences such as Interspeech. The recognition that both performance and availability of speech technologies is sharply unequal between different language communities, has further encouraged a focus on "inclusion" as highlighted in the special theme of this conference.

In this paper we want to draw attention to the motivations behind research on accent variation at Interspeech and the way researchers define and discuss "accented speech". Surveying papers published at Interspeech between 2004 and 2022, we notice that terms like "accent", "accented" and "non-native" are frequently used in under-specified ways that could hinder interpretability and reproducibility of research results. Drawing on a sociolinguistic perspective, we also make the case that some of the applications resulting from research on "accented speech", and the way researchers talk and write about accents could harm the very communities who are the intended benefactors of language technology development. We'd like to encourage researchers working on speech to think carefully about their research motivations and how that motivation affect what aspect of language variation really matters to their research question.

2. What's an accent?

In linguistics teaching and research, *accent* is often distinguished from *dialect*: *accent* describes pronunciation (*segmental* and *suprasegmental* phonology), while *dialect* also encompasses syntax and lexicon, (see e.g., Trudgill in the Encyclopedia for Language and Linguistics [1, 14] and Crystal in the Dictionary of Linguistics and Phonetics [2, 3]).

In everyday language, both *dialect* and *accent* are generally only used to describe some, non-standard varieties. Some speakers might be described as "having an accent" (implying that other people "don't have an accent"). In anglophone linguistics, "accent" and "dialect" are used as neutral descriptors – language varieties might differ in terms of phonetics, phonology, lexicon and syntax and they might differ in *social status* (depending largely on who speaks them) but they are all considered equally complex and rule-governed. As a result, **all varieties, including the "standard variety" could be described as dialects** and **everyone has an accent** [3].

Linguistic variation perceived as accent variation is often tied to the identity of a speaker or speaker group, in particular in terms of geography and social class [2, 3]. For example, in the context of British Englishes[1], "Received Pronunciation" (RP) is an accent associated in particular with upper class speakers, due to its use by upper class speakers and transmission in private schools [4, 5]. Other British English accents[2] are very strongly associated with particular regions, cities or areas, like Liverpool ("Scouse"), Manchester ("Mancunian"), Newcastle ("Geordie"), Glasgow ("Glaswegian") and (East) London (e.g., "Cockney" and "Multicultural London English"). Varieties associated with (post-)industrial areas like those, also retain strong associations with a lower socioeconomic status (similarly to how RP is associated with the upper class). In the context of the UK specifically, these associations still matter as some accents are considered to have "higher status" which can result in linguistic discrimination [6, 7, 8]. Studies also show that second language speakers and those perceived to have a "foreign accent" are affected by language-based discrimination, for example in employment [9, 10]. Second language speakers, especially racial or ethnic minorities, are often framed as "deficient" speakers who *lack* linguistic skills which speakers of the standard variety have [11, 12, 13].

To summarise, we can draw on Agha [4], who argues that the term *accent* is "neither very precise nor free of ideological distortion" (p 232). Firstly, *accent* often "implicitly presupposes a baseline against which some sound patterns — but not others — are focally perceived as deviant, foregrounded accents" [4, 232]. Secondly, accents do not *just* describe sound patterns in isolation but are inherently linked to a specific group of social identities [4]. Finally, accents are usually discussed as intrinsic features of a speaker (or their speech) which are either present or absent: some people don't have an accent, others do [4]. The reality is more complicated as the geographic or social descriptions of an accent depend on the listener's identity [4].

---

[1]For an excellent introduction to this issue in the context of the United States, see Lippi-Green [3]

[2]Most regional and/or social varieties also differ to some extent in lexicon and syntax, making them *dialects*. However, many people frequently apply their "native" accent to the "standard dialect", adapting only the phonology, not syntax and lexicon. In sociolinguistics, the term "variety" is often used to encompass languages, accents and dialects.

3. Methods

We analysed all 94 papers returned by the search term "accented" in the ISCA archive. To understand the characteristics of "accent" research in Interspeech, we also provide a cursory analysis of the 319 papers published on "accent" in this period.[3]

We manually categorised how the word "accent" or "accented" was applied in the abstract and/or introduction of the paper each paper: prosodic prominence (e.g., "pitch accent")[4], first language varieties (e.g., "native accent"), second language varieties (e.g., "foreign accent"), or methods relating to first or second language varieties (e.g., "multi-accent").[5] We then qualitatively analysed how the papers discuss "accent" and "accentedness". Specifically, we look at how accents, speakers and listeners are described and what motivations researchers provide for researching accents in speech technologies.[6]

4. Who has an accent?

As shown in Table 1, half of papers about "accent" published between 2004 and 2022 focus primarily on prosodic prominence – we won't discuss those further in this paper. Of the remaining 161 papers, less than a third are specifically about first language varieties and speakers (L1), with the rest explicitly addressing second language speakers and varieties (L2) or theories and methods concerning accent variation more broadly (L1-or-L2). Conversely, less than a third of papers using the term "accented" discuss prosody, with the plurality focusing on second language speakers and varieties.

Search term	L1	L2	L1-or-L2	Prosody	NA
"accent"	47	65	49	154	2
"accented"	15	37	14	28	0

Table 1: *Distribution of Interspeech papers by topic. "Accent" is most frequently used in the context of prosody, but "accented" is most frequently applied to L2 speakers and L2 varieties.*

Of the 66 papers discussing accent variation, most provide some descriptions of relevant accents. The level of detail in these descriptions varies widely, with some just naming relevant varieties or corpora (n=19), while others provide specific phonetic characteristics of the variety and/or some demographic details about the speakers and listeners involved (n=40). It is notable that most papers focussing on specific L1 or L2 varieties use abstract terms like "foreign accent" or "accented" in the paper title and abstract. Only 12 papers specifically name the relevant accent in the title, and another 11 mention the language (but not the accent). This approach emphasises the broad applicability of findings or methods much of the research aims for, where specific varieties are meant to serve as examples. This can be appropriate if the methods or findings truly generalise

beyond those specific varieties. In some cases generalisations about "foreign accents", do, however, reinforce the idea that all "accented speech" is very similar, which as both linguistic and speech technology literature shows, is not the case.

Describing speakers or language use as "accented" as in "accented speech" or "accented English" also implies the existence of "unaccented" speech or speakers. This is an explicit assumption in some Interspeech papers which refer to "unaccented" or "non-accented" speech or speakers, or speakers who have "no accent" (as discussed below). From a (socio)linguistic perspective these labels are not particularly meaningful, if we assume that all speech is characterised by an accent of some kind. In the "lay" context discussed in the introduction, "no accent" or "unaccented" is a way of referring to the "unmarked" variety, usually the standard variety. This may be quite difficult to interpret for readers unfamiliar with the sociolinguistic context, and, as discussed below, study participants may also differ in the way they interpret these terms.

5. Who are the speakers and listeners?

As Cheng et al. [15] highlight in the context of psycholinguistics, the vagueness of "non-native speaker" impedes effective study design, efficient recruitment of participants, clear interpretation of results, and, ultimately, reproducibility. Who is considered a "native speaker" varies between researchers and, importantly in self-reporting studies, among speakers themselves [15]. There is also huge variability between "non-native" speakers. Baese-Berk et al [16] note that there are some "common aspects of non-native speech" across different target and first languages, such as generally slower speech rate compared to L1 speakers and specific target language features which can be challenging for learners with a range of different backgrounds (e.g., two features of English: voiced stops in word-final position or vowel reduction in unstressed syllables) (p. 3).

However, most models of second language acquisition (grounded in empirical studies) posit that the phonology (system of speech sounds) and articulatory settings (the way speakers use their vocal tract habitually) of learner's "first" language has important effects on how they perceive and produce sounds in any additional languages [16]. As the strong interest in "accentedness" in the Interspeech literature evidences, L2 speakers who share the same first language still vary widely in their spoken language production. Some individual differences like habitual speech rate appear to carry across languages [17]. Furthermore, speakers also acquire sociolinguistic variation in their L2, depending on where, when and how they acquire and speak it [18, 19, 20]. To complicate this even further, the context of the speech recording such as task (e.g., reading/conversation), style (formal/informal), topic (e.g., topics which are or are not emotive like work, family, memories), relationship to the interlocutor and accent of the interlocutor have all been shown to affect how people speak in their L1 and L2 [16, 21].

Grouping together speakers with different linguistic backgrounds – both in terms of L1 and in terms of their exposure and use of L2 thus risks obscuring a lot of variation. As [16] note, perception crucially depends on the listener, not just the speaker. Listener expectations and (local) context such as the order in which stimuli of different speakers are presented, the degree of familiarity of listeners with different varieties and lexical frequency all affect perception tasks like accent classification or accentedness ratings [16, 22]. It is therefore particularly notable that many studies focusing on speech perception provide little or no description of speaker or listener demographics. Of

---

[3]The searches were conducted using the the "paper" search function in the ISCA archive. In February 2023, a search for "accented Interspeech" returned 96 results, with 94 available full papers."accent Interspeech" in the ISCA archive returned 320 results, with 319 available full papers published between 2004 and 2022.

[4]In linguistics, the term "accented" most frequently refers to prosodic prominence, i.e., lexical or pitch accent, as in "accented syllable" [2, 3].

[5]Papers investigating prosody in first or second language accents were categorised as L1 or L2 rather than prosody. 2 papers did not discuss any sense of the word "accent" and were categorised as NA.

[6]To avoid singling out specific authors, direct quotes are presented without attribution (following e.g., [14]).

the 17 perception studies, 7 only mention the L1 and gender of the speakers and 9 only mention L1 and gender of the listeners.

6. Why do we research accents?

The plurality of papers on "accented" speech focus on automatic speech recognition (n=27). Other popular topics are perception studies and phonetic description of different varieties (n=21) with fewer studies focusing primarily on language identification or speech synthesis.

A particularly interesting aspect of many perception studies are "accentedness ratings", which are employed in 15 papers. In these studies, listeners provide evaluations of second language speakers' accents (these ratings are not used for L1 speakers). As discussed above, the level of detail in the description of these listeners differs, but 14 studies confirm at least that they are native speakers and provide a gender distribution (one paper only mentions "human raters"), while some highlight relevant details such as familiarity with other varieties or residential history.

Most of these studies employ a scale ranging from "no accent" to "strong" or "heavy" accent. Two studies include scales of "no foreign accent" to "strong foreign accent". Three studies instead ask listeners to categorise short audio clips as "foreign" or "native". In this way almost all of the "accentedness rating" research explicitly invokes the notion of "unaccented speech" or "unaccented speakers". The distinction between "foreign" and "native" is also particularly complicated in pluricentric, global languages like English which have a larger number of very different "native" varieties and many multilingual "native" speakers. One study employing accentedness rating makes the implicit hierarchy within different varieties explicit by asking listeners to describe speakers as "native: the speaker sounds native (e.g., US, UK, Australian)" or "non-native: the speaker sounds like a learner of English (e.g., Korean, Japanese, Philippine)", which was, for the purposes of that study distinguished from different "degrees" of "Indian accent" of speakers who "sound Indian": "subtle", "clear", "pronounced" and "very thick".

While there are different motivations for these studies including understanding human language processing, pronunciation assessment is a common theme. As shown in the quote below (drawn from the aforementioned study focussed at "quantification of Indian accent"), at the most extreme end, this research can frame "accents" as "inappropriate":

- "To be successful in this industry, there is an increased demand for employers to be able to detect the heaviness of an accent so that they can assign employees to appropriate job categories, or give them additional training to refine their accents as appropriate for their jobs."

In addition to framing some ways of speaking as (in)appropriate for specific jobs (in this case, in a customer-facing call centre), it suggest "training" as a kind of remedy to this linguistic deficiency. The pressure placed in particular on migrants and workers in and from the Global South to participate in "accent reduction training" is well documented [12, 23]. It is embedded in wider discourses around "appropriate" or "professional" speech, in which "accent" or language is often used to stand in for race and where linguistic discrimination is inextricably linked to racism [11, 8].

A small number of studies focus on "accent reduction" from a technical perspective by applying "accent conversion" and on "speech error detection" using ASR. Both of these approaches are primarily motivated through use in second language teaching. Conceptually, this too relies on a notion of a "target pronunciation" or "correct pronunciation". While it is certainly the case that many learners of an additional language want to avoid miscommunication, not all pronunciation variation or even pronunciation "errors" lead to miscommunication (especially among human interlocutors who can often easily recover intended meaning by accessing the wider linguistic and non-linguistic context). Statements like the ones presented below reinforce notions of "one correct pronunciation", generalise across the extremely heterogeneous group of L2 speakers and "accents" as an impediment to (an undefined notion of) intelligibility "typical" (only) to L2 speakers.

- "Second-language (L2) English learners typically present accents and mispronunciations, which highly impact their intelligibility in practical communication."
- "Correct pronunciation is known to be the most difficult part to acquire for (native or non-native) language learners."
- "We focus on two major aspects of foreign accents: mispronunciations and improper prosody (rhythm, phonemes duration, and pauses)."
- "The goal of automatic pronunciation evaluation is to build an automatic system which can measure the quality of pronunciation given input speech."

These statements concerning "correct" pronunciation or "pronunciation quality" likely appear innocuous to most readers, including many linguists. Notions of "language proficiency" and "accentedness" as evaluated in comparison to some "ideal" or "prototypical" "native speaker" and the importance of standard varieties as linguistic targets for L1 and L2 speakers are deeply ingrained in language teaching [24]. While acknowledging that these notions, as well as models of "target pronunciations" can be useful for learners to avoid miscommunication and feel confident in their L2, it is also important to note that not all speakers orient towards "native speakers" [20]. Furthermore, as is particularly obvious in the context of (migrant) workers in international anglophone settings, "accent targets" are often imposed externally (e.g., by an employer) rather than freely chosen by learners [12, 10, 23].

7. Conclusions

Against this background, we encourage researchers working on accent variation, especially in the context of speech technology, to think carefully about the underlying assumptions and motivations of their research, and develop and apply nuanced approaches to variation in the context of specific research domain.

For instance, drawing on an ill-defined, or undefined notion of "accent" or "foreign accent" or "non-native speaker" risks erasing important variation in way that makes it much harder to solve real research problems like sharp differences in ASR performance for different varieties. While it is difficult to generalise about ASR performance about "foreign accents" it is possible to identify performance differences between more narrowly defined varieties of the same language [25, 26]. Close examination of the language variation which "triggers" speech recognition errors could further be used to improve systems [27, 28, 29].

Similarly, while it makes sense to keep descriptions of speakers, listeners and corpora brief in light of page limits at venues like Interspeech, we recommend thinking carefully about these description and providing any details required to replicate the study or interpret results. Describing the variety or accent with relevant phonological and/or social details is often appropriate. One option could be to include standardised lan-

guage variety tags (e.g., BCP-47) as recommended for natural language processing [30].

Being very specific in perception study design also aids reproducibility. "Accentedness" ratings, for instance, are not necessarily correlated with "intelligibility" [16]. Alternatively, some studies investigating intelligibility ask listeners to write down or re-speak what they heard in addition to or instead of "rating" speakers.

Finally, accents are just as much about identity as they are about pronunciation. The way we speak is always shaped by who we are and how we want to be perceived. On the one hand, that means that the extent to which we can definitively "label" different accents is limited, and that is important to be very clear about how we go about naming speakers and their varieties. On the other hand, the connection between identity and accent also means that while all accents are "equally valid" from a linguistic perspective, they are not all "equally valued" in society.

8. Acknowledgements

9. References

[1] P. Trudgill, "Accent," in *Encyclopedia of Language Linguistics (Second Edition)*, second edition ed., K. Brown, Ed. Elsevier, 2006, p. 14. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B0080448542015066

[2] D. Crystal, *Dictionary of Linguistics and Phonetics*. Wiley Sons, Incorporated, John, 2009.

[3] R. Lippi-Green, *English with an accent language, ideology, and discrimination in the United States*, 2nd ed. Routledge, 2012.

[4] A. Agha, "The social life of cultural value," *Language & Communication*, vol. 23, no. 3, pp. 231–273, 2003.

[5] A. H. Fabricius, "Social change, linguistic change and sociolinguistic change in received pronunciation," in *Sociolinguistics in England*, N. Braber and S. Jansen, Eds. Palgrave Macmillan UK, 2018, pp. 35–66.

[6] D. Sharma, E. Levon, and Y. Ye, "50 years of British accent bias: Stability and lifespan change in attitudes to accents," *English World-Wide*, vol. 43, pp. 135 – 166, 2022.

[7] E. Levon, D. Sharma, D. J. L. Watt, A. Cardoso, and Y. Ye, "Accent bias and perceptions of professional competence in england," *Journal of English Linguistics*, vol. 49, no. 4, pp. 355–388, 2021.

[8] J. T. Craft, K. E. Wright, R. E. Weissler, and R. M. Queen, "Language and Discrimination: Generating Meaning, Perceiving Identities, and Discriminating Outcomes," *Annual Review of Linguistics*, vol. 6, no. 1, pp. 389–407, 2020.

[9] A. R. Timming, "The effect of foreign accent on employability: a study of the aural dimensions of aesthetic labour in customer-facing and non-customer-facing jobs," *Work, Employment and Society*, vol. 31, no. 3, pp. 409–428, 2016-04, publisher: SAGE Publications.

[10] V. A. Ramjattan, "Racializing the problem of and solution to foreign accent in business," *Applied Linguistics Review*, vol. 13, no. 4, pp. 527–544, 2019.

[11] J. Rosa and N. Flores, "Unsettling race and language: Toward a raciolinguistic perspective," *Language in Society*, vol. 46, no. 5, pp. 621–647, 2017.

[12] V. A. Ramjattan, "Accenting racism in labour migration," *Annual Review of Applied Linguistics*, vol. 42, pp. 87–92, 2022.

[13] D. Wright and G. Brookes, "'This is England, speak English!': a corpus-assisted critical study of language ideologies in the right-leaning British press," *Critical Discourse Studies*, vol. 16, no. 1, pp. 56–83, 2018-08.

[14] R. K. Moore and L. Skidmore, "On the Use/Misuse of the Term 'Phoneme'," in *Proc. Interspeech 2019*, 2019, pp. 2340–2344.

[15] L. S. P. Cheng, D. Burgess, N. Vernooij, C. Solís-Barroso, A. McDermott, and S. Namboodiripad, "The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures," *Frontiers in Psychology*, vol. 12, 2021.

[16] M. M. Baese-Berk, D. J. McLaughlin, and K. B. McGowan, "Perception of non-native speech," *Language and Linguistics Compass*, vol. 14, no. 7, p. e12375, 2020.

[17] A. R. Bradlow, M. Kim, and M. Blasingame, "Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 886–899, 2017.

[18] L. Hall-Lew and Z. Elliott, "Production of FACE and GOAT by Slovak and Czech immigrants in Edinburgh," in *The 18th International Conference of the Phonetic Sciences*, 2015.

[19] M. Meyerhoff and E. Schleef, "Variation, contact and social indexicality in the acquisition of (ING) by teenage migrants," *Journal of Sociolinguistics*, vol. 16, no. 3, pp. 398–416, 2012.

[20] C. Nance, W. McLeod, B. O'Rourke, and S. Dunmore, "Identity, accent aim, and motivation in second language users: New Scottish Gaelic speakers' use of phonetic variation," *Journal of Sociolinguistics*, vol. 20, no. 2, pp. 164–191, 2016-04.

[21] W. Labov, "Some principles of linguistic methodology," *Language in Society*, vol. 1, no. 1, pp. 97–120, 1972. [Online]. Available: http://www.jstor.org/stable/4166672

[22] K. B. McGowan and A. M. Babel, "Perceiving isn't believing: Divergence in levels of sociolinguistic awareness," *Language in Society*, vol. 49, no. 2, pp. 231–256, 2019-10.

[23] C. Cowie, "The accents of outsourcing: the meanings of "neutral" in the Indian call centre industry," *World Englishes*, vol. 26, no. 3, pp. 316–330, 2007-08.

[24] I. Cushing and J. Snell, "The (white) ears of Ofsted: A raciolinguistic perspective on the listening practices of the schools inspectorate," *Language in Society*, pp. 1–24, 2022.

[25] N. Markl, "Language variation and algorithmic bias: Understanding algorithmic bias in British English automatic speech recognition," in *2022 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: Association for Computing Machinery, 2022, p. 521–534.

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

[27] M. P. Y. Chan, J. Choe, A. Li, Y. Chen, X. Gao, and N. Holliday, "Training and typological bias in ASR performance for world englishes," in *Interspeech 2022*. ISCA, 2022.

[28] J. Choe, Y. Chen, M. P. Y. Chan, A. Li, X. Gao, and N. Holliday, "Language-specific effects on automatic speech recognition errors for world englishes," in *Proceedings of the 29th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, 2022, pp. 7177–7186.

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

[30] E. M. Bender and B. Friedman, "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science," *Transactions of the Association for Computational Linguistics*, vol. 6, pp. 587–604, 2018.