L'analisi linguistica e letteraria XXXIII (2025) 33–56
ISSN: 1122-1917

Assessing Smartphone Speech Recognition Across Diverse English Accents: A Preliminary Study[1]

Claudia Soria, Rosalba Nodari, Silvia Calamai
Consiglio Nazionale delle Ricerche, Istituto di Linguistica Computazionale "A. Zampolli", Università di Siena
claudia.soria@cnr.it, rosalba.nodari@unisi.it, silvia.calamai@unisi.it

Received 28 April 2025; Accepted July 2025; Published online December 2025

This study examines the performance of a smartphone-based automatic speech recognition (ASR) system when processing diverse English accents. With the increasing reliance on voice-activated artificial intelligence in daily tasks, ensuring equitable ASR performance across linguistic varieties is critical. Using audio data from the CIRCE project corpus, we assess recognition accuracy for eleven English accents selected according to Kachru's three-circle model (Inner, Outer, and Expanding Circle varieties). Findings highlight disparities in recognition performance and suggest that ASR models exhibit a bias favoring American English (AmE). The study underscores the need for enhanced ASR inclusivity and diversification of training data.

Keywords: Automatic Speech Recognition, Smartphone, English Accents, Sociophonetics

1. *Introduction*

Accents are socially charged and ideologically filtered. Far from being neutral objects, they build social *personae* also in the digital and technological domain, where voice plays a relevant role. Accent is a fundamental dimension of linguistic variation, typically influenced by a speaker's geographical, social, or linguistic background. It involves phonetic and phonological variations such as vowel quality, consonant articulation, intonation, and rhythm, without necessarily affecting grammar or vocabulary (Ladefoged, Johnson 2015; Trudgill 2000; Lippi-Green 2012). Whether arising from second-language acquisition or from regional variation within a native language, accents serve as salient markers of identity, social background, and linguistic experience. In the context of second language (L2) speakers, the degree of foreign accent has often been linked to age of acquisition, quantity and qual-

[1] This study was supported by the project CIRCE (Counteracting Accent Discrimination Practices in Education, 2022-1-IT02-KA220-SCH-000087602), a three-year project funded by Erasmus+. The funding sources had no role in study design, data collection/analysis/interpretation, writing, and decision to submit. The three authors jointly developed the research and the design of the studies and collaboratively edited the entire paper. For academic purposes, the authors' responsibilities are the following. First author: conceptualization, methodology, investigation, writing, formal analysis, visualization; second author: data collection, conceptualization, investigation; third author: supervision, project coordination and administration, funding acquisition.

DOI: 10.69117/ALL.2025.3.04

ity of input, and, in some cases, personal traits and attitudes toward the target language and its associated culture (Flege 1995; Piske et al. 2001). The issue is particularly relevant in the case of a global language such as English: in today's world, where multilingualism is the norm rather than the exception, people who speak English as a second or foreign language far outnumber those who speak it as their first language. Decolonial perspectives and multilingual pedagogies further underscore the importance of affirming the legitimacy of Englishes different from British and American and of shifting attention toward more peripheral varieties.

Kachru's foundational sociolinguistic framework, commonly known as the "Three Circles model" (1985), classifies global Englishes based on their historical trajectories, sociopolitical functions, and cultural status. Although critiqued for reinforcing hierarchies and failing to capture the dynamic nature of global English (Galloway, Rose 2015), Kachru's concentric model still remains influential in World English studies. The model differentiates between three varieties: Inner Circle varieties, where English is spoken as a first language (e.g., the United States, the United Kingdom); Outer Circle varieties, where English plays a second-language role in multilingual societies (e.g., India, Nigeria); and Expanding Circle varieties, where English is primarily used as a foreign language and lacks official institutional status (e.g., China, Italy, Turkey)[2]. Among first language (L1) English speakers, regional and ethnolectal variation plays a crucial role in shaping pronunciation and perceived speaker identity. Studies have shown that varieties such as African American English (AAE), Multicultural London English (MLE), and Southern American English (SAE) can be subject to social evaluation, being usually judged as less prestigious forms (Lippi-Green 2012). These social evaluations are usually mirrored in technological bias (Holliday, Villareal 2020; Lawrence 2021; Hofmann et al. 2024) as these accents, though phonologically systematic and locally accepted, often diverge from the standardized norms encoded in speech technologies, leading to disparities in recognition accuracy and user experience (Koenecke et al. 2020; Choe et al. 2022). Thus, while research in sociophonetics and applied linguistics emphasizes that accents present meaningful variation that cannot be dismissed as "deviation" from a standard, but must be understood in relation to broader sociocultural structures, in Automatic Speech Recognition (ASR) research they are typically treated as such (Prinos et al. 2024). Critically, ASR systems are not neutral to variation in speaker accent or dialect, although any human-machine interaction occurs through an accented version of a specific language in real-world usage.

In recent years, smartphones have become indispensable tools for communication, productivity, and entertainment (Nawaz 2024). Their growing integration with artificial intelligence (AI) (Goggin 2025) has made ASR a key component of human-machine interaction (Li et al. 2022), enabling functionalities such as voice assistants, real-time transcription, and hands-free control. Voice-activated virtual assistants like Google Assistant, Apple Siri, Amazon Alexa, and Samsung Bixby are now widely used, with increasingly

[2] In the remainder of this paper, Inner, Outer and Expanding varieties will be identified with the labels "INN", "OUT", and "EXP".

broad linguistic coverage (Jaber et al. 2024; Palanica, Fossat 2021). For instance, Google Assistant supports over 40 languages[3], Apple's Siri 26[4], and Amazon Alexa 9[5].

Despite remarkable technological advances and vendor claims of enhanced user experience[6], real-world usage reveals important limitations, particularly in how accurately these assistants recognize speech from users speaking with non-standard accents (Rìo et al. 2023; Lawrence 2021). A substantial body of research has shown that these systems systematically underperform when processing speech from speakers with so-called "non-standard" or "regional" English accents (Tatman, Kasten 2017; Tatman 2017; Choe et al. 2022; Michel et al. 2025)[7].

Although some providers offer localized, dialect-specific models (e.g., Siri supports nine English dialects), empirical studies continue to show that ASR performance is skewed toward dominant varieties, particularly Standard American English, while speakers of non-standard or underrepresented accents are disadvantaged (Jain et al. 2018; Koenecke et al. 2020; O'Neill, Carson-Berndsen 2023).

Tatman (2017), for instance, already demonstrated significant disparities in YouTube's auto-generated captions across gender and dialect groups. Harris et al. (2024) further confirmed this pattern by evaluating leading ASR models on speakers of Standard American English versus three minority dialects, finding consistently lower transcription accuracy for the latter. Similarly, L2 speech continues to present challenges for ASR. Studies have reported systematic biases, with higher Word Error Rates (WERs) for L2 speakers (Cámbara et al. 2021; Prinos et al. 2024). Tadimeti et al. (2022) assessed various commercial and academic ASR systems and found that performance degraded considerably for non-American accents, with absolute accuracy gaps ranging from 2% to 12% and relative differences of 16% to 49%.

A growing body of literature has established that ASR accuracy is tightly linked to the composition of training data. Systems trained predominantly on standard varieties tend to generalize poorly to regional, L2, or less common L1 accents. Although recent work has introduced promising techniques for mitigating these biases, such as geographically stratified corpora and accent-aware modeling (Maison, Estève 2023; Do et al. 2024; Deng et al. 2021; Najafian, Russell 2020), multiple studies still report significant disparities in ASR performance. Kuhn et al. (2024) identified consistently higher WERs for L2 English

[3] https://venturebeat.com/ai/google-assistant-can-now-interpret-44-languages-on-smartphones/ (last accessed July 23, 2025).

[4] https://en.wikipedia.org/wiki/Siri (last accessed July 23, 2025).

[5] https://en.wikipedia.org/wiki/AmazonAlexa (last accessed July 23, 2025).

[6] https://www.wired.com/story/siri-ios-11-update-improvement-voice/ (last accessed July 23, 2025) and https://aws.amazon.com/pm/transcribe/ (last accessed July 23, 2025).

[7] We are aware that the traditional dichotomies, such as *standard* vs. *non-standard* or *native* vs. *non-native* varieties, are insufficient to fully capture the diverse functional and social dimensions of English varieties spoken worldwide, as one reviewer critically observed. At the same time, this is not the venue to address such a complex and multi-layered debate; we therefore refer the reader to Berruto's seminal essay *Miserie e grandezze dello standard. Considerazioni sulla nozione di standard in linguistica e sociolinguistica* (Berruto 2007), which offers an in-depth discussion of the notion of "standard" from both linguistic and sociolinguistic perspectives.

speakers and Graham and Roll (2024) observed that OpenAI's Whisper achieved better accuracy on L1 English accents compared to L2 ones. Similarly, Choe et al. (2022) and Hollands et al. (2022) emphasized that performance disparities persist across demographic and linguistic lines, particularly between L1 and L2 English users.

These findings underscore a persistent mismatch between technological capabilities and real-world linguistic diversity. ASR systems continue to perform best on L1 or standardized English speech, often failing to meet the needs of users who speak with different accents or come from linguistically marginalized communities (Wenzel et al. 2023; Brewer et al. 2023; Kulkarni et al. 2024).

In global contexts such as healthcare, education, and remote work, where English is often used as a *lingua franca*, this discrepancy in ASR performance can lead to communication inefficiencies and reduced accessibility for a substantial segment of the global population. Patients may experience breakdowns in communication during medical interactions (Miner et al. 2020; Tobin et al. 2024); students may struggle with inaccurate ASR-based captions or note-taking tools (Kuhn et al. 2024; Tatman 2017); job seekers may face subtle forms of linguistic bias in AI-powered screening or productivity tools (Hoffmann et al. 2024). What can be considered merely as a technical underperformance may thus risk reinforcing existing linguistic hierarchies and contributing to a broader form of marginalization and linguistic inequality (Sun Eidsheim 2023; Drożdżowicz, Peled 2024; Michel et al. 2025), whereby speakers of less widely spoken or less prestigious accents are made to feel invisible or illegible to mainstream technologies.

Against this backdrop, the present study aims to assess whether Apple Siri, one of the most widely deployed ASR systems, exhibits accent-based performance disparities based on users' accents. By contributing to the literature on algorithmic bias in speech technologies, this research aims to raise awareness about persistent linguistic inequities in digital communication and encourage developers to design more inclusive systems. The remainder of this paper is organized as follows. Section 2 outlines the methodological framework, including the rationale for accent selection, data sources, and experimental procedures. Section 3 presents and interprets the results, combining quantitative performance metrics with a qualitative error typology to capture both accuracy and functional impact across accent groups. Section 4 discusses the limitations of the study and outlines directions for future research, particularly in relation to dataset expansion, multi-platform comparisons, and alternative evaluation metrics. Finally, Section 5 concludes by synthesizing the main findings, reflecting on their implications for linguistic equity in speech technologies, and proposing strategies for developing more inclusive and accent-aware ASR systems.

2. *Methodology*

2.1 Background and Objectives

Addressing the sociolinguistic complexity of English is particularly relevant in the context of ASR, where performance disparities may reflect not only linguistic distance from "standard" varieties but also structural biases in model training and data selection. To frame the

analysis of ASR performance of Apple iPhones across English varieties, this study draws on Kachru's Three Circles Model (1985), allowing us to move beyond a simplistic and problematic native/non-native dichotomy and to foreground the social, geopolitical, and ideological dimensions of accent in English speech technologies. By incorporating Kachru's model in the analysis, we aim to situate accents within broader global hierarchies of language prestige, technological representation, and access. In doing so, we intend to foreground the uneven socio-technical landscape in which ASR systems operate and to lay the groundwork for a more equitable evaluation of recognition performance across diverse English-speaking communities.

This study evaluates the automatic speech recognition performance of Apple iPhones across a sample of accents belonging to Inner, Outer, and Expanding Circles varieties. In particular, we mean to:

1. Test whether localized ASR model variants yield better performance when aligned with a speaker's accent or regional variety.
2. Compare the performance of localized models with that of the generic "English (default)" model on the same speech inputs.
3. Explore whether performance differences reflect systematic biases affecting certain accent groups (e.g., INN vs. OUT/EXP).

We define *localization* here as the adaptation of ASR models to specific regional or national varieties of English (e.g., "English [UK]," "English [Australia]"), as offered in the iOS settings[8]. Although not introduced earlier, this concept plays a central role in this study and is based on the premise that linguistic localization improves ASR accuracy by aligning system behavior with users' pronunciation, vocabulary, and regional conventions. The choice of a specific localized model (or *variant*) can be influenced by multiple factors, including the speaker's accent, geographic and cultural context, and the effectiveness of speech recognition for personal use. Users often select a variant that aligns with their pronunciation to improve system comprehension, especially if they have a distinct accent. Geographical location also plays a role, as localised models are typically optimized for interactions with regional services, and lexical differences between English varieties may further influence this choice, reducing the risk of misinterpretation. Additionally, compatibility with other applications, such as voice-controlled devices, may determine which variant is most functional. Personal preference is another consideration, particularly for bilingual individuals or those with exposure to multiple varieties of English.

Building on these observations, we hypothesize that aligning one's accent with a corresponding ASR variant should improve recognition accuracy (RQ1). We also hypothesize that, if geographic location is a determining factor in ASR model optimization (as is often the case for localized ASR systems designed to interact with regional services), then choosing a variant based on one's location, even when it does not match the speaker's accent, should yield recognition performance at least comparable to that of the default generic model (RQ2). Finally, we seek to investigate whether systematic performance disparities

[8] As of 2024, they were the USA, the UK, Australia, Canada, India, Japan, New Zealand, Singapore, South Africa.

emerge among different accent groups, particularly in relation to speakers of Inner vs. Outer or Expanding Circle varieties of English (RQ3). Rather than assuming bias *a priori*, we use the term here to refer to any measurable performance asymmetry that may be considered unfair or disproportionate across speaker categories.

2.2 Data

This study draws on audio samples from the CIRCE corpus, which features read-text recordings from speakers of diverse English accents. The dataset includes speech from male and female speakers across four Inner and seven Outer and Expanding Circle varieties of English.

To ensure comparability across accent groups, all participants read the same 74-word reference passage (see below). This controlled elicitation protocol minimizes content-related variability and allows for a consistent assessment of recognition accuracy:

> Food products often travel large distances to reach a store. An orange, for example, might travel by truck from a farm to a packaging plant. It might then be exported by plane to another country, where it reaches a store and then, finally, a consumer. That is why more and more people claim that transporting foods over large distances is harmful to the environment and choose to eat food products from small, local farms.

The dataset includes the following accents:

- Inner Circle: Standard American English (AmE), African American English (AAE), Standard British English (BrE), Multicultural London English (MLE).
- Outer Circle: Indian (InE), Nigerian (NiE).
- Expanding Circle: Bosnian (BoE), Italian (ItE), Turkish (TuE), Ukrainian (UkE), Chinese (ChE).

The eleven accents were chosen to reflect both dominant and underrepresented varieties, including ethnically and regionally marked dialects (e.g., AAE, MLE) often overlooked in ASR evaluation. While resource constraints played a role in accent inclusion (especially for less-represented varieties such as Bosnian or Ukrainian), these voices were not arbitrarily chosen. All recordings were sourced from the CIRCE corpus, which pre-classified speakers based on detailed background metadata. Accent representativeness was verified by trained phoneticians affiliated with the corpus project and confirmed by the authors through manual inspection of the metadata and speech samples.

Each accent group includes one male and one female voice, yielding a total of 22 audio clips. Each clip is approximately 32 seconds long on average. While detailed sociolinguistic variables (e.g., age of onset, time in English-speaking contexts) were not made available through the corpus, the metadata does include speaker gender, age range, and accent self-identification. These were used to screen the recordings for eligibility and balance.

Recordings were made using standardized procedures defined by the corpus compilers. These included:

- Quiet environments (minimal background noise);
- Consistent microphone setup across speakers;
- Controlled reading task with visual prompts.

As all recordings came from the same corpus, channel-related acoustic variability was minimized.

2.3 Tools and Procedure

As discussed in Section 1, although Siri is not the most widely used or technically advanced ASR system (Palanica, Fossat 2021), it remains one of the most accessible, being pre-installed on all Apple smartphones. With over 2.3 billion iPhones sold globally since 2007[9], Siri has significant market penetration. Moreover, among the major commercial voice assistants, Siri offers the broadest support for English regional variants (Lawrence 2021), making it a suitable platform for evaluating accent-sensitive ASR performance in real-world consumer contexts.

All testing was conducted in March 2024 using the iOS 17.4 operating system on an iPhone 13 running the latest version of Siri at the time of testing. Recordings were transcribed via Siri's on-device voice dictation feature embedded in the Apple Notes app. Audio clips were played from a desktop computer at a consistent volume using built-in speakers. The iPhone was positioned ~15–20 cm from the speaker output to simulate realistic smartphone usage scenarios such as voice messaging or dictation. Each of the 22 audio clips (11 accents × 2 speakers) was transcribed three times with each of the nine available English-language Siri variants[10], resulting in a total of 594 transcriptions. The ASR output for each clip was compared against the reference passage, and the Word Error Rate (WER) was calculated for each passage and as an average of the three passages for each accent and voice. WER is a standard metric for ASR evaluation, defined as:

$$\mathrm{WER} = (S + D + I) / N$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions, and N is the total number of words in the reference transcript (74).

WER is computed using Levenshtein distance and provides a quantitative measure of ASR accuracy. While widely used, WER is limited in that it treats all word-level errors equally, regardless of semantic weight or grammatical function (Szymanski et al. 2020; Tobin et al. 2022; Coro et al. 2025).

To interpret ASR performance for applications such as dictation and smartphone voice assistants, we adopted the following scale, updated to reflect recent usability research (Harrington 2023):

- Excellent (WER < 5%): human-like performance; suitable for dictation and high-stakes usage.
- Good (5% ≤ WER < 7%): minor errors occur but do not significantly hinder usability. Suitable for general-purpose dictation and everyday assistant use.
- Acceptable (7% ≤ WER < 10%): noticeable errors; may require user corrections or re-phrasing. Usable for casual voice commands, limited dictation.

[9] https://en.wikipedia.org/wiki/IPhone (last accessed July 22, 2025).

[10] As tested in 2024.

– Marginal (10% ≤ WER < 15%): errors may impair user experience; only acceptable in constrained tasks or low-stakes use.
– Unacceptable (WER ≥ 15%): not viable for dictation or voice assistant deployment.

This performance categorization enables an interpretable, user-oriented assessment of Siri's ASR behavior across English variants and accent types.

3. *Results and Discussion*

3.1 General Overview

Table 1 presents the aggregate performance of each Siri English localized variant across all accents. Results confirm that no model performs at an Excellent or even Acceptable level overall. The UK, New Zealand, and Canada models yield the lowest average Word Error Rates (WERs), but still exceed the 10% threshold typically considered acceptable for dictation and virtual assistant applications. Models localized for Japan, India, and Singapore perform markedly worse, with mean WERs exceeding 20% and in some cases approaching 60%, making them impractical for reliable use in general-purpose ASR tasks.

Table 1 - *General performance of Siri English variants (across all accents)*

Model	*Mean*	*Median*
UK	11.69	11.50
AU	15.36	13.50
US	14.06	12.20
CA	13.45	12.20
JP	58.80	60.80
IN	23.55	19.60
NZ	13.32	11.50
SI	23.39	16.90
ZA	15.14	13.50

Table 2 displays performance averaged over all Siri models for each of the 11 tested accents. The only accent with a mean WER below 5% is Standard American English (AmE), which meets the "Excellent" threshold. African American English (AAE) and Indian English (InE) follow, both within the "Acceptable" range. Other Inner Circle accents, including Multicultural London English (MLE) and British English (BrE), show higher error rates, in some cases worse than several Outer and Expanding Circle varieties. Notably, Chinese (ChE) and Nigerian (NiE) English exhibit better recognition scores than BrE, MLE, or Italian English (ItE). These findings challenge the assumption that Inner Circle varieties always perform better than Outer and Expanding Circle varieties.

Table 2 - *Average performance per accent (aggregated across Siri variants)*

Accent	*Mean*	*Median*
AAE	10.71	9.50
AmE	3.96	2.70
BrE	23.72	19.60
MLE	17.46	15.55
BoE	28.53	23.00
ChE	10.15	13.50
InE	8.35	8.10
ItE	20.25	16.20
NiE	15.24	12.85
TuE	17.71	16.20
UkE	22.61	23.00

To further assess the reliability of model performance, we also report standard deviation and computed 95% confidence intervals for each accent (see Table 3). Accents such as InE, AmE, and AAE exhibit low variability, suggesting consistent recognition across Siri models. In contrast, accents like Bosnian English (BoE), BrE, and ItE show wide fluctuations, reflecting inconsistent model behavior.

Table 3 - *Accent performance consistency across models: Mean WER, Standard Deviation, and 95% Confidence Intervals*

Accent	*Mean WER (%)*	*StDev*	***95% CI (±)***
InE	8.35	2.51	**±1.44**
AmE	3.96	3.11	**±1.78**
AAE	10.71	4.58	**±2.63**
ChE	10.15	5.59	**±3.21**
UkE	22.61	5.69	**±3.26**
TuE	17.71	9.03	**±5.18**
MLE	17.46	9.44	**±5.42**
NiE	15.24	11.16	**±6.40**
ItE	20.25	15.15	**±8.69**
BoE	28.53	17.44	**±10.01**
BrE	23.72	24.21	**±13.90**

The observed trend of high accuracy coupled with low variability for AmE and InE likely reflects a confluence of three factors:

1. Extensive representation in training data: both accents are well represented in commercial ASR corpora, resulting in consistent performance across models (Koenecke et al. 2020; Blodgett et al. 2020).

2. Phonetic alignment with dominant ASR norms: their phonological patterns approximate so-called standard varieties, reducing model error during phoneme decoding (Lippmann 1997; Adank et al. 2004).
3. Industry prioritization: these accents are commercially valuable and prioritized in system optimization, especially for customer-facing applications such as call centers and smart assistants (Tatman 2017; Veliche et al. 2024).

Conversely, accents like BoE, BrE, and ItE not only suffer from higher mean WERs, but also from high variability: this may indicate that models are inconsistent in their ability to process these speech patterns. This variability suggests gaps in training coverage, phonetic dissimilarity to dominant varieties, or poor generalization capacity.

Finally, the presence of high WER variability (especially for Inner Circle varieties such as MLE and AAE and most Expanding Circle varieties) underscores the importance of evaluating both performance and consistency when assessing ASR fairness.

3.2 Accent-specific Model Performance

Tables 4 and 5 present a detailed breakdown of ASR recognition performance across the eleven English accents and nine ASR variants. The data show substantial differences both between models and across accents.

Table 4 - *WER (%) per accent and ASR variant, mean values, aggregated female and male speakers*

Accent	*AU*	*CA*	*IN*	*JP*	*NZ*	*SI*	*UK*	*US*	*ZA*
AAE	10.58	9.47	15.33	54.73	9.02	11.97	9.03	10.17	10.15
AmE	2.93	2.95	8.58	31.77	3.18	7.23	2.07	1.58	3.18
BrE	20.95	13.77	44.62	72.73	17.80	45.72	16.92	14.20	15.77
MLE	14.65	12.85	35.12	83.82	13.30	22.52	11.28	15.10	14.88
BoE	29.07	34.93	23.43	73.42	20.95	30.42	16.02	31.30	42.13
ChE	9.25	9.70	15.32	48.20	8.57	10.83	9.47	9.25	8.80
InE	8.58	6.55	7.90	33.10	7.68	9.70	9.23	8.13	9.02
ItE	16.22	14.65	22.75	74.33	17.55	47.30	13.75	14.43	15.32
NiE	11.05	10.38	31.07	67.35	11.05	28.13	7.68	9.72	12.85
TuE	20.27	12.17	24.78	61.03	14.88	20.25	12.62	19.83	16.90
UkE	25.45	20.50	30.18	46.83	22.52	23.18	20.50	20.97	17.57

Table 5 - *WER (%) per accent and ASR variant, median values, aggregated female and male speakers*

Accent	*AU*	*CA*	*IN*	*JP*	*NZ*	*SI*	*UK*	*US*	*ZA*
AAE	10.80	9.45	14.20	62.15	9.50	12.20	8.80	10.15	10.15
AmE	2.70	2.70	8.15	29.75	3.40	7.45	1.40	2.05	3.40
BrE	20.25	12.85	41.25	76.35	16.90	45.25	15.55	14.90	15.55
MLE	16.20	11.50	37.15	82.45	13.55	21.60	10.85	12.85	14.20
BoE	18.90	25.70	20.95	76.35	22.30	27.70	16.25	29.05	35.15

Accent	*AU*	*CA*	*IN*	*JP*	*NZ*	*SI*	*UK*	*US*	*ZA*
ChE	8.80	8.80	14.90	45.95	8.80	11.50	10.15	9.45	8.80
InE	9.50	6.10	6.80	33.10	7.45	10.15	8.80	7.45	8.80
ItE	16.20	14.85	23.65	75.00	16.20	43.90	12.85	14.20	15.55
NiE	10.85	12.20	29.70	66.25	11.50	31.75	6.80	8.80	12.85
TuE	19.60	11.50	23.65	60.80	14.20	20.25	12.15	21.65	16.25
UkE	25.00	20.95	29.75	46.60	21.60	23.65	20.30	21.65	16.90

Inner Circle varieties generally yield lower Word Error Rates (WERs) than Outer and Expanding Circle accents, with AmE consistently achieving the best recognition scores. The Canadian and US models, in particular, demonstrate high robustness, showing strong performance across multiple accents. In contrast, the Japanese model remains a consistent outlier, with WERs significantly above the defined threshold of usability across nearly all accents.

A closer look at median values in Table 5 reveals that fewer than half of all model-accent combinations (48 out of 99) fall within an acceptable range of performance. While some models, particularly CA, NZ, UK, and US, exceed acceptability on 3 to 4 accents each, others, such as SI and IN, underperform across most accents. These findings reinforce results from Section 3.1, showing that only a small subset of model-accent combinations meet the standards expected for practical deployment.

From the accent perspective, AmE is recognized above the acceptability threshold by 8 out of 9 models (excluding the JP model), with performance reaching the "Excellent" category in most cases. ChE and InE follow, with 5 models each demonstrating at least acceptable performance. Recognition of AAE is moderate, with 3 models reaching acceptability. At the other end of the spectrum, BoE, Turkish (TuE), Ukrainian (UkE), and ItE are never recognized above the threshold by any model, highlighting persistent limitations in current ASR systems for less dominant Expanding Circle varieties.

Table 6 synthesizes this information categorically. Among Inner Circle varieties, AmE achieves the most consistent performance, while BrE and MLE perform poorly across nearly all models. The failure of MLE to be recognized acceptably by any model is especially noteworthy and aligns with earlier concerns about the limited sociolinguistic coverage in training data. BrE, despite its historical prominence, also fails to achieve usability thresholds, suggesting either misalignment with acoustic training norms or insufficient model tuning.

Table 6 - *WER acceptability per accent and ASR variant, median values, aggregated female and male speakers*

Accent	*AU*	*CA*	*IN*	*JP*	*NZ*	*SI*	*UK*	*US*	*ZA*
AAE	M	**A**	M	U	**A**	M	**A**	M	M
AmE	**E**	**E**	**A**	U	**E**	**A**	**E**	**E**	**E**
BrE	U	M	U	U	U	U	U	M	U
MLE	U	M	U	U	M	U	M	M	M

Accent	*AU*	*CA*	*IN*	*JP*	*NZ*	*SI*	*UK*	*US*	*ZA*
BoE	U	U	U	U	U	U	U	U	U
ChE	**A**	**A**	M	U	**A**	M	M	**A**	**A**
InE	**A**	**G**	**G**	U	**A**	M	**A**	**A**	**A**
ItE	U	M	U	U	U	U	M	M	U
NiE	M	M	U	U	M	U	**G**	**A**	M
TuE	U	M	U	U	M	U	M	U	U
UkE	U	U	U	U	U	U	U	U	U

Among Outer and Expanding Circle varieties, ChE and InE are the best-performing, especially when paired with more robust models (e.g., CA, UK, and US). These accents may benefit from broader global exposure and inclusion in multilingual datasets. In contrast, recognition of BoE and UkE is particularly weak, with high WERs and zero instances of acceptable recognition. This underperformance highlights a broader issue of typological and geographical imbalance in ASR training corpora.

Importantly, only AmE reaches the "Excellent" level in most cases. Some accents do reach "Good" or "Acceptable" levels under specific conditions: for instance, InE achieves "Good" scores with CA and IN models, while ChE is recognized acceptably by AU, CA, and US models.

Overall, these results emphasize:

- the dominance of AmE-centric optimization across systems;
- a systematic performance gap between high-resource and low-resource accents;
- the limitations of current models in reliably supporting linguistic diversity, particularly in underrepresented Expanding Circle varieties.

3.2.1 Performance of Localized ASR Models on Corresponding Accents

To evaluate whether localized ASR variants are optimized for the accents they ostensibly support, we examined three models (US, UK, and IN) on their regionally aligned accents. These include AmE and AAE for the US model; BrE and MLE for the UK model; and Indian English for the IN model.

The results suggest mixed outcomes. For AmE, both the US and UK models perform excellently, with the UK model showing slightly superior performance. Interestingly, the UK model also outperforms the US model for AAE, indicating that localization does not always equate to superior performance on aligned accents. For BrE, the UK model is outperformed by both the Canadian and American models, while for MLE, the UK model does perform best, though still below the acceptability threshold. Regarding Indian English, both the IN and CA models yield strong results, with the CA model showing a slight edge.

These findings show that geographic alignment alone does not guarantee optimal performance. While some localized models perform well, other non-localized systems (especially CA) often exhibit greater cross-accent generalizability. This suggests that training data diversity and ASR model design may matter more than regional tagging *per se*.

3.2.2 Accent–Model Matching: Does Localization Benefit the Speaker?

A more user-centric question is whether selecting the localized ASR model variant that corresponds to the speaker's country of residence improves recognition for their accent. This analysis is especially relevant for speakers of Outer and Expanding Circle varieties of English, whose accents are not typically tied to a single national context.

Our findings show that some localized models offer recognition advantages for certain accents. For instance, AAE is more accurately recognized by the Canada, New Zealand, and UK models than by the US model. Similarly, BrE is better recognized by CA and US models than by the UK variant. InE achieves optimal scores with both the IN and CA models. ChE benefits most from the ZA model.

However, these performance gains are uneven and inconsistent. For AmE, recognition is uniformly excellent across all models. By contrast, MLE, BoE, ItE, TuE, and UkE consistently underperform across models, with no localized system offering meaningful improvement. This pattern suggests that localization mainly benefits accents already well represented in training data, with diminishing returns for those less frequently encountered.

Moreover, the assumption underlying localized ASR variants that a user's accent corresponds to their geographic location does not hold for many speakers of Expanding Circle varieties of English. Accents such as ItE or TuE may be spoken globally, independent of a single national territory. A speaker of Italian-accented English may reside in France, Germany, or the UK and use English in international or digital domains. In such cases, selecting a "local" model provides no inherent benefit. Instead, these users often default to standard varieties (e.g., UK or US English), which may or may not be better tuned to their phonetic features.

This points to a structural limitation in the current localization paradigm. While useful for geographically bound Inner Circle varieties (e.g., Scottish English in Scotland), it offers limited utility for transnational speakers. Future ASR development may need to shift from geographically localized models to accent-sensitive personalization strategies that adapt to users' speech patterns dynamically, regardless of location.

3.3 Implications and Interpretation

The results of this study confirm substantial variation in ASR performance across English accents, with clear disparities that reflect underlying biases in model training and optimization. Standard American English and Indian English consistently achieve the highest levels of recognition accuracy and the lowest standard deviation, suggesting both strong representation in training data and alignment with the phonetic norms prioritized in ASR development.

In contrast, other varieties, including some L1 accents such as British English and Multicultural London English, display higher WERs and greater variability, indicating limited exposure during training and phonetic divergence from the dominant modeling standards. Surprisingly, several Outer and Expanding Circle varieties (e.g., Indian, Chinese) outperform certain Inner Circle accents, revealing that performance is shaped not strictly by nativeness but by broader sociotechnical dynamics, including commercial priorities and data availability.

These findings underscore the entanglement between linguistic ideologies and technological design. From a sociolinguistic perspective, particularly within the framework of globalization (Blommaert 2010), ASR technologies appear to reinforce new global hierarchies of English. It would seem, thus, that Kachru's (1985) concentric-circle model, once useful in describing the geopolitical diffusion of English, could be increasingly inadequate in accounting for technological mediation. While only one inner-circle variety (AmE) consistently performs well, Indian English (a historically outer-circle variety) now rivals it in ASR accuracy. This shift reflects the growing influence of countries like India and China, not only in terms of user base but also as pivotal centers of speech technology development and linguistic modelling (see, for instance, Javed et al. 2022; Fang et al. 2022).

As such, ASR performance may no longer mirror traditional sociolinguistic prestige or native/non-native distinctions. Instead, it increasingly aligns with the infrastructural and commercial realities of the tech sector, where dominant training paradigms are shaped by regions with large markets, technical expertise, or strategic importance. This dynamic raises important questions about the sociotechnical power shaping what counts as "recognizable" English.

3.4 Qualitative Error Typology: Toward Linguistically Informed ASR Evaluation

While Word Error Rate (WER) is a useful benchmark, it is widely recognized that it offers no insight into the qualitative nature of errors or their functional impact. To address this limitation, we conducted a focused typological analysis of ASR outputs, following recent frameworks developed for linguistically meaningful ASR evaluation (e.g., Papadopoulou et al., Zaretskaya, Mitkov 2021; Meripo, Konam 2022).

A subset of the transcriptions generated by Siri was selected for qualitative analysis, specifically those produced by the US model, which achieved the highest average performance across English varieties. For each speaker and accent, we selected the transcription corresponding to the median Word Error Rate (WER). These transcriptions were subsequently tokenized, automatically annotated for part-of-speech, and manually coded according to error type, part-of-speech category involved, and impact on intelligibility. The classification scheme adopted was adapted from Wirth and Peinl (2022), with modifications to suit the aims of the present study:

- Edit operation type (*deletion*, *insertion* or *substitution*): these tags categorize automatic transcription errors based on the three fundamental operations used in calculating the Word Error Rate.
- Grammatical category: this level refers to the category of part-of-speech affected by the deletion, insertion or substitution. Tags are *punctuation* (such as deletion or insertion of a punctuation mark), *grammar* (i.e. substitution of "a" with "the" or of "travels" with "travel"), lexicon ("stone" vs. "store"), and *lexico-grammar*, when the error involved the substitution of a lexical item with a functional one and vice-versa ("large" vs. "by"; "by" vs. "bike").
- Impact on comprehensibility of resulting text: *negligible* (such as in omission of punctuation marks), *minor*, and *major* (such as substitution of lexical items affecting semantic content, such as "explored" instead of "exported" or "helpful" instead of "harmful").

Table 7 illustrates the percentage of Edit operation type of errors, showing that the type of error alone cannot be linked with better or worse recognition by ASR[11].

Table 7 - *Distribution of ASR errors by Edit operation type*

Accent	*WER*	*DEL*	*INS*	*SUB*
AmE	2.70	69.23	7.69	23.08
AAE	8.80	64.71	11.76	23.53
BrE	14.20	71.79	5.13	23.08
MLE	14.20	58.97	10.26	30.77
InE	7.45	56.67	10.00	33.33
NiE	9.50	60.61	12.12	27.27
BoE	31.05	76.19	6.35	17.46
ChE	9.50	50.00	12.50	37.50
ItE	14.20	53.66	9.76	36.59
TuE	18.90	65.31	6.12	28.57
UkE	20.95	49.06	7.55	43.40

Table 8 presents the PoS affected by the errors introduced by the ASR system.

Table 8 - *Error typology based on linguistic category*

Accent	*WER*	*PUN*	*GRA*	*GRA-LEX*	*LEX*
AmE	2.70	76.92	0.00	33.33	15.38
AAE	8.80	61.76	23.53	0.00	14.71
BrE	14.20	48.72	35.90	11.11	10.26
MLE	14.20	46.15	28.21	0.00	25.64
InE	7.45	60.00	23.33	30.00	6.67
NiE	9.50	57.58	18.18	55.56	9.09
BoE	31.05	28.57	31.75	0.00	39.68
ChE	9.50	56.25	21.88	8.33	18.75
ItE	14.20	46.34	36.59	13.33	12.20
TuE	18.90	42.86	26.53	7.14	26.53
UkE	20.95	37.74	35.85	26.09	15.09

Accent-specific analysis of error distributions reveals notable variation in how different linguistic categories contribute to ASR performance, as measured by Word Error Rate (WER). Accents with the highest WER such as BoE and UkE present elevated rates of lexical and grammatical errors, indicating that failures in recognizing content and struc-

[11] The values for WER reported in this and the two following tables refer to the average between the two transcriptions considered per each accent.

ture-bearing words significantly undermine transcription accuracy. In contrast, accents with lower WERs, such as AmE, InE, and AAE, tend to exhibit fewer lexical and grammatical errors, with a larger proportion of punctuation errors, which are typically less detrimental to intelligibility. Interestingly, the "GRA/LEX" category, which captures substitution errors between grammatical and lexical items (e.g., replacing a noun with a preposition), shows sporadic distribution across accents. While these hybrid errors are infrequent overall, their occurrence, particularly in Indian and American English, suggests complex misalignments between syntactic roles that can affect parsing and meaning reconstruction, even when WER is low. Notably, high WER values are consistently associated with a prevalence of lexical errors, reinforcing the critical role of content-word fidelity in ASR accuracy. The analysis suggests that lexical misrecognition is a more damaging error type than grammatical or punctuation-based deviations.

A Pearson correlation test was thus conducted to assess whether the four dimensions were statistically correlated with WER. Results show a strong negative association between WER and Punctuation ($r = -0.96$) and a strong positive correlation between WER and lexical items ($r = 0.73$). The correlation with the "GRA" error category is moderately strong and positive ($r = 0.67$), whereas the correlation with the GRA/LEX category is weak and negative ($r = -0.44$). These results suggest that punctuation errors, although frequent, do not directly contribute to the WER; in fact, a high number of punctuation errors can coexist with good ASR performance. Conversely, WER tends to increase substantially with the number of lexical errors. This is expected, as WER reflects substitutions, deletions, or insertions of words, which are often lexical in nature. Grammatical errors also contribute to WER, though their impact appears slightly lower than that of lexical errors. In conclusion, WER appears to be more sensitive to lexical than to grammatical errors, suggesting that it is not neutral to error type. This differential impact points to the need for error typologies in ASR evaluation that distinguish between the functional weight of error types, particularly in multilingual and multi-accented contexts.

Table 9 presents the percentage distribution of the different types of error impact introduced by the automatic speech recognition analysis.

Table 9 - *Distribution of ASR errors by semantic impact across accents*

Accent	*WER*	*NEG*	*MIN*	*MAJ*
AmE	2.70	0.00	84.62	15.38
AAE	8.80	58.82	17.65	23.53
BrE	14.20	48.72	28.21	23.08
MLE	14.20	7.69	46.15	46.15
InE	7.45	43.33	30.00	26.67
NiE	9.50	48.48	21.21	30.30
BoE	31.05	14.29	25.40	60.32
ChE	9.50	56.25	15.63	28.13

Accent	*WER*	*NEG*	*MIN*	*MAJ*
ItE	14.20	31.71	39.02	29.27
TuE	18.90	44.90	10.20	44.90
UkE	20.95	37.74	9.43	52.83

A detailed analysis of error severity across accents reveals a strong positive correlation between WER and the proportion of major errors (MAJ), with a coefficient of 0.90. This suggests that accents with higher WERs are not only prone to more frequent errors but also to more impactful ones that significantly affect semantic interpretation. Conversely, WER is moderately negatively correlated with minor errors (MIN, $r = -0.46$), indicating that systems performing better tend to produce more superficial, less disruptive mistakes. The correlation between WER and negligible errors (NEG) is weak ($r = -0.15$), suggesting limited relevance. These findings reinforce the need to complement quantitative metrics like WER with qualitative measures of error impact, especially when evaluating ASR usability and fairness across accent groups.

4. *Limitations and Future Work*

While this study provides valuable insights into ASR performance across diverse English accents, several limitations must be acknowledged.

First, the dataset includes only two speakers per accent (one male, one female), limiting the generalizability of findings and precluding intra-accent analysis. Expanding the sample to include more speakers across age, region, and sociolinguistic profiles would enhance representativeness and enable more robust statistical evaluation (Blodgett et al. 2020). We also acknowledge that the dataset does not fully cover the global range of L2 Englishes. However, the diversity achieved here enables a meaningful first-step comparison of Inner, Outer, and Expanding Circle accents in consumer-grade ASR systems.

Second, the use of read speech ensures lexical and syntactic consistency across samples, but it does not fully replicate the variability and disfluency of spontaneous speech. Future work should incorporate conversational data or spontaneous speech to better assess ASR performance in real-world contexts (Kuhn et al. 2024).

Third, the analysis relies exclusively on Apple's Siri, limiting the scope to one ASR system. Including additional commercial and open-source models such as Google Assistant, Amazon Alexa, or OpenAI Whisper would allow for comparative analysis across platforms and help determine whether observed patterns are Siri-specific or reflect broader industry trends (DiChristofano et al. 2024; Graham, Roll 2024).

Fourth, Word Error Rate (WER), though a standard metric, provides a limited view of transcription quality. Even though our sample analysis appears to suggest a correlation between WER and both the impact and type of errors, we are aware that WER treats all errors equally, does not account for semantic fidelity, and offers no insight into user impact. Future research should combine WER with alternative metrics (e.g., Semantic Error

Rate, entity recognition accuracy) and qualitative error analyses to capture usability and real-world implications more fully (Szymański et al. 2020; Coro et al. 2025).

Finally, the limited number of observations precludes formal statistical testing. Future work should calculate confidence intervals around central estimates and increase the dataset size to enable significance testing (e.g., ANOVA or Kruskal-Wallis tests) to determine whether observed differences are statistically robust.

Despite these limitations, the study presents strong preliminary evidence of systemic disparities in ASR performance. These findings underscore the urgent need for more inclusive, geographically and phonologically diverse training data, and for ASR evaluation frameworks that center equity alongside accuracy.

5. *Conclusions*

This study contributes to ongoing research on the evaluation of automatic speech recognition systems by examining how accent variation affects ASR performance on widely available smartphone technology. By comparing the recognition accuracy of multiple localized English variants of Apple's Siri across a controlled set of various English accents, the study offers empirical insight into accent-based disparities in ASR performance.

Our quantitative analysis revealed clear disparities in the performance of smartphone-based ASR systems across English accents. Standard American English consistently outperformed all other varieties, with Indian and Chinese English also achieving relatively strong recognition scores. In contrast, many Expanding Circle accents, such as Bosnian, Turkish, Ukrainian and Italian, were systematically underrecognized. Notably, even some Inner Circle accents like British English and Multicultural London English yielded high error rates, showing that belonging to Inner Circle varieties alone does not necessarily guarantee accurate recognition. These results align with prior research showing that ASR systems tend to perform best on speech varieties most represented in their training data and support the growing call for inclusive data curation and model testing frameworks that account for sociolinguistic diversity in speech. The stronger performance of American and Indian English may reflect the linguistic preferences embedded in commercial development pipelines and market-oriented ASR optimization.

The qualitative error analysis provides further insight into the disparities in ASR performance, showing that WER is more sensitive to lexical errors than to grammatical ones. While punctuation errors were frequent, they had little effect on WER or overall intelligibility. In contrast, lexical substitutions were often semantically misleading, particularly when nouns were replaced with other content words or when a content word was replaced with a function word (and vice versa). These findings confirm that the type and communicative impact of an error matter as much as its mere occurrence and suggest the limitations of relying solely on WER as a performance metric.

A novel contribution of this study lies in its examination of regional ASR variants and their interaction with both L1 and L2 accents. Although geographic alignment between localized ASR model variant and speaker accent did not always lead to improved recogni-

tion, in some cases (e.g., for Indian or Chinese English) localized or regionally proximate models offered modest benefits. However, the study also found that such advantages were unevenly distributed and did not extend to less widely represented Expanding Circle varieties, indicating that current localization strategies may be insufficient to address the full range of linguistic diversity encountered in global English usage.

Whilst we are aware of the methodological limitations of this study (as detailed in section 4), we believe it offers methodological and conceptual contributions to the study of linguistic equity in speech technology. In particular, the use of a standardized passage across all accents, coupled with a multi-model evaluation approach, enabled consistent comparisons while highlighting underexamined dynamics in localized ASR performance. Such an approach can lend itself to providing a clearer understanding of where and why recognition systems succeed or fail, and how these outcomes intersect with the sociolinguistic hierarchies of English in a globalized world. Crucially, these findings suggest that disparities in ASR performance do not entirely mirror the broader hierarchical structure of English varieties described in Kachru's Three Circles model, reflecting changing global patterns of linguistic prestige, technological representation, and access.

Developing more inclusive ASR systems is not merely a matter of technological optimization but one of linguistic equity. The risk of accent bias is not only that it impairs user experience, but that it may reinforce existing social inequalities in digital interaction. As literature attentive to the underlying mechanisms of algorithms highlights, the choice not to worry about algorithms constitutes a privilege that today is no longer acceptable. If a person conforms to societal norms and also speaks with an accent deemed socially desirable, the algorithms are likely to function even more favorably for him (Noble 2018).

It is precisely in this context that sociolinguistics must assert its significance and role. When non-technological issues are addressed exclusively through technological lenses, technology experts often have the final say, frequently developing solutions that are more complex than the original problems they sought to resolve, while their effectiveness is almost impossible to evaluate (Morozov 2011). Therefore, we hope that this study will contribute to the growing body of research advocating for accent-aware ASR evaluation and the systematic redesign of speech technologies to support linguistic diversity.

References

Adank, Patti, Roeland van Hout, Roel Smits. 2004. "An Acoustic Description of the Vowels of Northern and Southern Standard Dutch." *Journal of the Acoustical Society of America* 116 (3): 1729–1738. https://doi.org/10.1121/1.1779271.

Berruto, Gaetano. 2007. "Miserie e grandezze dello standard. Considerazioni sulla nozione di standard in linguistica e sociolinguistica." In *Standard e non standard tra scelta e norma*, a cura di Paolo Molinelli, 13–41. Roma: Il Calamo.

Blodgett, Su Lin, Solon Barocas, Hal Daumé III, Hanna Wallach. 2020. "Language (Technology) Is Power: A Critical Survey of 'Bias' in NLP." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky, Joyce Chai, Natalie

Schluter, Joeal Tetreault, 5454–5476. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.48.

Blommaert, Jan. 2010. *The Sociolinguistics of Globalization*. Cambridge: Cambridge University Press.

Brewer, Robin N., Christina Harrington, Courtney Heldreth. 2023. "Envisioning Equitable Speech Technologies for Black Older Adults." In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 379–388. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3593013.3594005.

Cámbara, Guillermo, Alex Peiró-Lilja, Mireia Farrús, Jordi Luque. 2021. "English Accent Accuracy Analysis in a State-of-the-Art Automatic Speech Recognition System." In *Proceedings of PaPE 2021 Workshop "From speech technology to big data phonetics and phonology*." https://doi.org/10.48550/arXiv.2105.05041.

Choe, June, Yiran Chen, May Pik Yu Chan, Aini Li, Xin Gao, Nicole Holliday. 2022. "Language-Specific Effects on Automatic Speech Recognition Errors for World Englishes." In *Proceedings of the 29th International Conference on Computational Linguistics*, edited by Nicoletta Calzolari et al., 7177–7186. International Committee on Computational Linguistics. https://aclanthology.org/2022.coling-1.628.

Coro, Gianpaolo, Francesco Cutugno, Loredana Schettino, Emilia Tanda, Alessandro Vietti, Vincenzo Norman Vitale. 2025. "Phoné: An Initiative to Develop a Dataset for the Automatic Recognition of Spoken Italian." *Oral Archives Journal* 1: 89–107. https://doi.org/10.36253/oar-3340.

Deng, Keqi, Songjun Cao, Long Ma. 2021. "Improving Accent Identification and Accented Speech Recognition under a Framework of Self-Supervised Learning." In *Proceedings of Interspeech 2021*, 1504–1508. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2021-1186.

DiChristofano, Alex, Henry Shuster, Shefali Chandra, Neal Patwari. 2024. "Performance Disparities between Accents in Automatic Speech Recognition." In *Proceedings of the AAAI Conference on Artificial Intelligence* 37 (13): 16200–16201. https://doi.org/10.1609/aaai.v37i13.26960.

Do, Cong-Thanh, Shuhei Imai, Rama Doddipatla, Thomas Hain. 2024. "Improving Accented Speech Recognition Using Data Augmentation Based on Unsupervised Text-to-Speech Synthesis." In *Proceedings of EUSIPCO 2024*, 136–140. European Association for Signal Processing. https://doi.org/10.48550/arXiv.2407.04047

Drożdżowicz, Anna, Yael Peled. 2024. "The Complexities of Linguistic Discrimination." *Philosophical Psychology* 37 (6): 1459–1482. https://doi.org/10.1080/09515089.2024.2307993.

Fang, Zheng, Ruiqing Zhang, Zhongjun He, Hua Wu, Yanan Cao. 2022. "Non-Autoregressive Chinese ASR Error Correction with Phonological Training." In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, edited by Marine Carpuat, Marie-Catherine de Marneffe, Ivan Vladimir Meza Ruiz, 5907–5917. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.432.

Flege, James E. 1995. "Second Language Speech Learning: Theory, Findings, and Problems." In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by Winifred Strange, 233–277. Timonium, MD: York Press.

Galloway, Nicola, Heath Rose. 2015. *Introducing Global Englishes*. London: Routledge.

Goggin, Gerard. 2025. "Mobile AI: Communication and Mobility After the Smartphone." *Communication and Change* 1 (1): 5. https://doi.org/10.1007/s44382-025-00002-3.

Graham, Calbert, Nathan Roll. 2024. "Evaluating OpenAI's Whisper ASR: Performance Analysis across Diverse Accents and Speaker Traits." *JASA Express Letters* 4 (2): 025206. https://doi.org/10.1121/10.0024876.

Harrington, Lauren. 2023. "Incorporating Automatic Speech Recognition Methods into the Transcription of Police-Suspect Interviews: Factors Affecting Automatic Performance." *Frontiers in Communication* 8. https://doi.org/10.3389/fcomm.2023.1165233.

Harris, Camille, Chijioke Mgbahurike, Neha Kumar, Diyi Yang. 2024. "Modeling Gender and Dialect Bias in Automatic Speech Recognition." In *Findings of the Association for Computational Linguistics: EMNLP 2024*, edited by Yaser Al-Onaizan, Mohit Bansal, Yun-Nung Chen, 15166–15184. Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.890.

Hofmann, Valentin, Pratyusha Ria Kalluri, Dan Jurafsky, Sharese King. 2024. "AI Generates Covertly Racist Decisions about People Based on Their Dialect." *Nature* 633 (8028): 147–154. https://doi.org/10.1038/s41586-024-07856-5.

Hollands, Sam, Daniel Blackburn, Heidi Christensen. 2022. "Evaluating the Performance of State-of-the-Art ASR Systems on Non-Native English Using Corpora with Extensive Language Background Variation." In *Proceedings of Interspeech 2022*, 3958–3962. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2022-10433.

Holliday, Nicole, Dan Villarreal. 2020. "Intonational Variation and Incrementality in Listener Judgments of Ethnicity." *Laboratory Phonology* 11 (1): 3. https://doi.org/10.5334/labphon.229.

Jaber, Razan, Sabrina Zhong, Sanna Kuoppamäki, Aida Hosseini, Iona Gessinger, Duncan P. Brumby, Benjamin R. Cowan, Donald Mcmillan. 2024. "Cooking with Agents: Designing Context-aware Voice Interaction". In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–13. Association for Computing Machinery. https://doi.org/10.1145/3613904.3642183.

Jain, Abhinav, Minali Upreti, Preethi Jyothi. 2018. "Improved Accented Speech Recognition Using Accent Embeddings and Multi-Task Learning." In *Proceedings of Interspeech 2018*, 2454–2458. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2018-1864.

Javed, Tahir, Sumanth Doddapaneni, Abhigyan Raman, Kaushal Santosh Bhogale, Gowtham Ramesh, Anoop Kunchukuttan, Pratyush Kumar, Mitesh M. Khapra. 2022. "Towards Building ASR Systems for the Next Billion Users". In *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (10): 10813–10821. Association for the Advancement of Artificial Intelligence. https://doi.org/10.1609/aaai.v36i10.21327.

Kachru, Braj B. 1985. "Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle". In *English in the World: Teaching and Learning the Language and Literatures,* edited by Randolph Quirk, H.G. Widdowson, 11–30. Cambridge: Cambridge University Press.

Koenecke, Allison, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, Sharad Goel. 2020. "Racial Disparities in Automated Speech Recognition." In *Proceedings of the National Academy of Sciences* 117 (14): 7684–7689. https://doi.org/10.1073/pnas.1915768117.

Kuhn, Korbinian, Verena Kersken, Benedikt Reuter, Niklas Egger, Gottfried Zimmermann. 2024. "Measuring the Accuracy of Automatic Speech Recognition Solutions." *ACM Transactions on Accessible Computing* 16 (4). https://doi.org/10.1145/3636513.

Kulkarni, Ajinkya, Atharva Kulkarni, Miguel Couceiro, Isabel Trancoso. 2024. "Unveiling Biases while Embracing Sustainability: Assessing the Dual Challenges of Automatic Speech Recognition Systems." In *Proceedings of Interspeech 2024*, 4628–4632. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2024-2494.

Ladefoged, Peter, Keith Johnson. 2015. *A Course in Phonetics*. 7th ed. Boston: Cengage Learning.

Lawrence, Halcyon M. 2021. "Siri disciplines." In *Your Computer Is on Fire*, edited by Thomas S. Mullaney, Benjamin Peters, Mar Hicks, Kavita Philip, 179–198. MIT Press.

Li, Bo, Tara N. Sainath, Ruoming Pang, Shuo-Yiin Chang, Qiumin Xu, Trevor Strohman, Vince Chen, Qiao Liang, Heguang Liu, Yanzhang He, Parisa Haghani, Sameer Bidichandani. 2022. "A Language Agnostic Multilingual Streaming On-Device ASR System." In *Proceedings of Interspeech* 2022, 3188–3192. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2022-10006.

Lippi-Green, Rosina. 2012. *English with an Accent: Language, Ideology, and Discrimination in the United States*. London: Routledge.

Lippmann, Richard P. 1997. "Speech Recognition by Machines and Humans." *Speech Communication* 22 (1): 1–15. https://doi.org/10.1016/S0167-6393(97)00021-6.

Maison, Lucas, Yannick Estève. 2023. "Improving Accented Speech Recognition with Multi-Domain Training." In *Proceedings of ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICASSP49357.2023.10096268.

Meripo, Nimshi Venkat, Sandeep Konam. 2022. "ASR Error Detection via Audio-Transcript Entailment." In *Proceedings of Interspeech 2022*, 3358–3362. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2022-11177.

Michel, Shira, Sufi Kaur, Sarah Elizabeth Gillespie, Jeffrey Gleason, Christo Wilson, Avijit Ghosh. 2025. "'It's Not a Representation of Me:' Examining Accent Bias and Digital Exclusion in Synthetic AI Voice Services." In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 228–45. Association for Computing Machinery. https://doi.org/10.1145/3715275.3732018.

Miner, Adam S., Albert Haque, Jason A. Fries, Scott L. Fleming, Denise E. Wilfley, G. Terence Wilson, Arnold Milstein, Dan Jurafsky, Bruce A. Arnow, W. Stewart Agras, Fei-Fei Li, Nigam H. Shah. 2020. "Assessing the Accuracy of Automatic Speech Recognition for Psychotherapy." *npj Digital Medicine* 3 (1): 82. https://doi.org/10.1038/s41746-020-0285-8.

Morozov, Evgeny. 2011. *The Net Delusion: The Dark Side of Internet Freedom*. New York: PublicAffairs. https://doi.org/10.1017/S1537592711004014.

Najafian, Maryam, Martin Russell. 2020. "Automatic Accent Identification as an Analytical Tool for Accent Robust Automatic Speech Recognition." *Speech Communication* 122: 44–55. https://doi.org/10.1016/j.specom.2020.05.003.

Nawaz, Saqib. 2024. "Distinguishing between Effectual, Ineffectual, and Problematic Smartphone Use: A Comprehensive Review and Conceptual Pathways Model for Future Research." *Computers in Human Behavior Reports* 14: 100424. https://doi.org/10.1016/j.chbr.2024.100424.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press. https://doi.org/10.2307/j.ctt1pwt9w5.

O'Neill, Emma, Julie Carson-Berndsen. 2023. "Investigating the Sensitivity of Automatic Speech Recognition Systems to Phonetic Variation in L2 Englishes." *University of Pennsylvania Working Papers in Linguistics* 29 (2): 109–118. https://repository.upenn.edu/handle/20.500.14332/58897.

Palanica, Adam, Yan Fossat. 2021. "Medication Name Comprehension of Intelligent Virtual Assistants: A Comparison of Amazon Alexa, Google Assistant, and Apple Siri Between 2019 and 2021." *Frontiers in Digital Health* 3: 669971. https://doi.org/10.3389/fdgth.2021.669971.

Papadopoulou, Martha Maria, Anna Zaretskaya, Ruslan Mitkov. 2021. "Benchmarking ASR Systems Based on Post-Editing Effort and Error Analysis." In *Proceedings of the Translation and Interpreting Technology Online Conference*, edited by Ruslan Mitkov, Vilelmini Sosoni, Julie Christine Giguère, Elena Murgolo, and Elizabeth Deysel, 199–207. INCOMA Ltd. https://aclanthology.org/2021.triton-1.23/.

Piske, Thorsten, Ian R.A. MacKay, James E. Flege. 2001. "Factors Affecting Degree of Foreign Accent in an L2: A Review." *Journal of Phonetics* 29 (2): 191–215. https://doi.org/10.1006/jpho.2001.0134.

Prinos, Kerri, Neal Patwari, Cathleen A. Power. 2024. "Speaking of Accent: A Content Analysis of Accent Misconceptions in ASR Research." In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1245–1254. Association for Computing Machinery. https://doi.org/10.1145/3630106.3658969.

Río, Miguel, Corey Miller, Ján Profant, Jennifer Drexler-Fox, Quinn Mcnamara, Nishchal Bhandari, Natalie Delworth, Ilya Pirkin, Migüel Jetté, Shipra Chandra, Hendoro Peter, Ryan Westerman. 2023. "Accents in Speech Recognition through the Lens of a World Englishes Evaluation Set." *Research in Language* 21: 225–244. https://doi.org/10.18778/1731-7533.21.3.02.

Szymański, Piotr, Piotr Żelasko, Mikolaj Morzy, Adrian Szymczak, Marzena Żyła-Hoppe, Joanna Banaszczak, Lukasz Augustyniak, Jan Mizgajski, Yishay Carmiel. 2020. "WER We Are and WER We Think We Are." In *Findings of the Association for Computational Linguistics: EMNLP 2020*, edited by Trevor Cohn, Yulan He, Yang Liu, 3290–3295. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.295.

Sun Eidsheim, Nina. 2023. "7. Rewriting Algorithms for Just Recognition: From Digital Aural Redlining to Accent Activism." In *Thinking with an Accent: Toward a New Object, Method, and Practice,* edited by Pooja Rangan, 134–150. Berkeley: University of California Press. https://doi.org/10.1525/9780520389748-012

Tadimeti, Divya, Kallirroi Georgila, David Traum. 2022. "Evaluation of Off-the-Shelf Speech Recognizers on Different Accents in a Dialogue Domain." In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, edited by Nicoletta Calzolari et al., 6001–6008. European Language Resources Association. https://aclanthology.org/2022.lrec-1.645.

Tatman, Rachael. 2017. "Gender and Dialect Bias in YouTube's Automatic Captions." In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, edited by Dirk Hovy, Shannon Spruit, Margaret Mitchell, Emily M. Bender, Michael Strube, and Hanna Wallach, 53–59. Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-1606.

Tatman, Rachael, Christopher Kasten. 2017. "Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and YouTube Automatic Captions." In *Proceedings of Interspeech 2017*, 934–938. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2017-1746.

Tobin, Jimmy, Qisheng Li, Subhashini Venugopalan, Katie Seaver, Richard Cave, Katrin Tomanek. 2022. "Assessing ASR Model Quality on Disordered Speech using BERTScore." In *Proceedings of the 1st Workshop on Speech for Social Good (S4SG)*, 26–30. International Speech Communication Association. https://doi.org/10.21437/S4SG.2022-6.

Tobin, Jimmy, Phillip Nelson, Bob MacDonald, Rus Heywood, Richard Cave, Katie Seaver, Antoine Desjardins, Pan-Pan Jiang, Jordan R. Green. 2024. "Automatic Speech Recognition of

Conversational Speech in Individuals with Disordered Speech." *Journal of Speech, Language, and Hearing Research* 67 (11): 4176–4185. https://doi.org/10.1044/2024_JSLHR-24-00045.

Trudgill, Peter. 2000. *Sociolinguistics: An Introduction to Language and Society*. 4th ed. London: Penguin Books.

Veliche, Irina-Elena, Zhuangqun Huang, Vineeth Ayyat Kochaniyan, Fuchun Peng, Ozlem Kalinli, Michael L. Seltzer. 2024. "Towards Measuring Fairness in Speech Recognition: Fair-Speech Dataset." In *Proceedings of Interspeech 2024*, 1385–1389. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2024-2273

Wenzel, Kimi, Nitya Devireddy, Cam Davison, Geoff Kaufman. 2023. "Can Voice Assistants Be Microaggressors? Cross-Race Psychological Responses to Failures of Automatic Speech Recognition." In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–14. Association for Computing Machinery. https://doi.org/10.1145/3544548.3581357.

Wirth, Joannes, Rene Peinl. 2022 "Automatic Speech Recognition in German: A Detailed Error Analysis." *2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, 1–8. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/COINS54846.2022.9854978.