Voice assistant comprehension of Cameroonian English: Technological linguistic imperialism in the digital age

Azane Charles, PhD
University of Buea
abimnui.azane@ubuea.cm

Abstract
This study investigates the accuracy of mainstream voice assistants (Siri, Google Assistant, and Alexa) in comprehending Cameroonian English, a variety spoken by over 8 million speakers. Through experimental phonetics and systematic error analysis involving 15 Cameroonian English speakers performing 2,250 voice commands across three platforms, the study demonstrates significantly lower comprehension rates for Cameroonian English (56.8% accuracy) compared to documented performance with Standard American and British English (95% accuracy). Phonological features including syllable-timed rhythm, consonant cluster simplification, and distinctive vowel realisations emerged as primary sources of recognition failure. Drawing on sociolinguistic theory, this research reveals how artificial intelligence technologies reproduce linguistic discrimination, creating digital exclusion for speakers of non-dominant English varieties. The paper argues that this constitutes technological linguistic imperialism that restricts equitable access to essential digital services. The study concludes with recommendations for developing inclusive automatic speech recognition systems and situates findings within broader conversations about linguistic justice in technology design.

Key words: *accent discrimination, automatic speech recognition, Cameroonian English, linguistic imperialism, technological bias, voice assistants, World Englishes*

Résumé

Cette étude examine la précision des assistants vocaux grand public (Siri, Google Assistant et Alexa) dans la compréhension de l'anglais camerounais, une variété parlée par plus de 8 millions de locuteurs. À travers la phonétique expérimentale et l'analyse systématique des erreurs impliquant 15 locuteurs d'anglais camerounais effectuant 2 250 commandes vocales sur trois plateformes, l'étude démontre des taux de compréhension significativement plus faibles pour l'anglais camerounais (56,8 % de précision) par rapport aux performances documentées avec l'anglais américain et britannique standard (95 % de précision). Les caractéristiques phonologiques incluant le rythme syllabique, la simplification des groupes consonantiques et les réalisations vocaliques distinctives sont apparues comme les principales sources d'échec de reconnaissance. S'appuyant sur la théorie sociolinguistique, cette recherche révèle comment les technologies d'intelligence artificielle reproduisent la discrimination linguistique, créant une exclusion numérique pour les locuteurs de variétés d'anglais non dominantes. L'article soutient que cela constitue un impérialisme linguistique technologique qui restreint l'accès équitable aux services numériques essentiels. L'étude conclut avec des recommandations pour développer des systèmes de reconnaissance automatique de la parole inclusifs et situe les résultats dans le cadre de conversations plus larges sur la justice linguistique dans la conception technologique.

Mots-clés: *Anglais Camerounais, assistants vocaux, biais technologique, discrimination accentuelle, impérialisme linguistique, reconnaissance automatique de la parole, variétés mondiales de l'anglais*

1. Introduction

The proliferation of voice-activated technologies has fundamentally transformed human-computer interaction, with voice assistants like Apple's Siri, Amazon's Alexa, and Google Assistant becoming ubiquitous interfaces for digital services. By 2024, an estimated 8.4 billion digital voice assistants were in use globally (Statista, 2024), mediating access to information, communication, smart home controls, and essential services. However, this technological revolution conceals a troubling inequity in that these systems demonstrate marked performance disparities across different English varieties, effectively privileging speakers of dominant accents while marginalising those who speak non-Western varieties of English.

This study examines voice assistant comprehension of Cameroonian English, an understudied but demographically significant variety spoken by over 8 million people in a bilingual nation where English serves as one of the two official languages alongside French. Cameroon's unique sociolinguistic landscape, shaped by colonial history, multilingualism, and contact between English and over 280 indigenous languages, has produced distinctive phonological, lexical, and prosodic features in its English variety (Kouega, 2007; Mbangwana, 2004). Despite its legitimacy as a recognised variety within the World Englishes paradigm (Kachru, 1985), Cameroonian English speakers report frequent frustration with voice-activated technologies that fail to understand their speech.

The implications extend beyond mere inconvenience. As digital services increasingly rely on voice interfaces, from banking and healthcare to education and emergency services, inability to effectively use these technologies creates digital exclusion that

disproportionately affects speakers of non-dominant English varieties. This exclusion intersects with existing inequalities in technology access, economic opportunity, and educational resources, compounding disadvantages for populations in the Global South.

This research addresses three interconnected objectives. First, the study quantifies the accuracy disparity in voice assistant comprehension of Cameroonian English compared to documented performance with dominant varieties. Second, the researcher identifies specific phonological features that trigger recognition failures. Third, the researcher analyses these findings through sociolinguistic frameworks, examining how automatic speech recognition (ASR) technologies function as gatekeeping mechanisms that enforce linguistic hierarchies in the digital sphere.

Research questions

This study addresses three primary research questions:

RQ1: How accurately do mainstream voice assistants comprehend commands issued in Cameroonian English compared to documented performance with Standard American and British English?

RQ2: Which specific phonological features of Cameroonian English correlate most strongly with voice assistant comprehension failures?

RQ3: How do Cameroonian English speakers experience and interpret voice assistant comprehension failures, and what do these experiences reveal about linguistic discrimination in digital technologies?

2. Theoretical framework and literature review

This section establishes the theoretical foundations for understanding voice assistant comprehension disparities as a form of linguistic discrimination. It synthesizes three interconnected bodies of scholarship: the World Englishes paradigm, which legitimizes non-dominant varieties; research on automatic speech recognition bias, which documents technological performance gaps; and critical perspectives on linguistic imperialism in digital technologies. Together, these frameworks illuminate how ASR systems perpetuate linguistic hierarchies that disadvantage speakers of varieties like Cameroonian English.

World Englishes and sociolinguistic variation

The World Englishes paradigm, established by Kachru (1985), recognises English as a pluricentric language with multiple legitimate varieties shaped by different historical, cultural, and linguistic contexts. Kachru's three-circle model positions Cameroonian English within the Outer Circle, which consists of territories where English functions as a second language with official status due to colonial history. However, sociolinguistic scholarship has critiqued this framework for its implicit hierarchisation, with Inner Circle varieties maintaining prestige while Outer and Expanding Circle varieties face delegitimisation (Pennycook, 2017).

Cameroonian English exhibits distinctive features across all linguistic levels. Phonologically, it demonstrates syllable-timed rhythm rather than the stress-timed rhythm characteristic of British and American English (Mbangwana, 2004). Speakers often simplify consonant clusters word-finally (Kouega, 2007), and vowel systems show systematic differences including monophthongisation of diphthongs influenced by substrate

languages from Niger-Congo and Afro-Asiatic families (Simo Bobda, 1994). These features reflect natural language contact processes and are systematic within the Cameroonian English speech community, yet they diverge from phonological norms encoded in ASR training data.

Critical sociolinguistics emphasises that linguistic variation intersects with power relations. Varieties associated with politically and economically dominant groups gain prestige, while those spoken by marginalised populations are stigmatised through language ideologies, and culturally situated beliefs about the nature and value of different ways of speaking. Standard language ideology, which positions certain varieties as inherently correct or superior, underpins much linguistic discrimination, including in technological contexts.

Automatic speech recognition and accent bias

ASR technology relies on machine learning models trained on massive speech corpora. However, these training datasets disproportionately represent speakers of dominant English varieties, particularly Standard American English. Studies have documented significant performance disparities across accent groups, with error rates increasing substantially for speakers of Scottish English, African American Vernacular English, and South Asian English compared to Standard American English speakers (Feng et al., 2021; Koenecke et al., 2020; Tatman, 2017). Research by Koenecke et al. (2020) for example, found that five leading ASR systems exhibited racial disparities, with average word error rates nearly twice as high for Black speakers compared to white speakers. Also, Tatman (2017) demonstrated that YouTube's automatic captioning system was significantly more likely to accurately transcribe male voices than female voices, with even

larger gaps for speakers with non-American accents. These findings suggest systematic bias rather than random error.

The technical origins of this bias lie in data collection practices. ASR models require thousands of hours of transcribed speech for training, and assembling such datasets favours easily accessible speaker populations in technology development centres which are primarily urban, filled with affluent speakers of dominant varieties in the United States and United Kingdom (Blodgett et al., 2020). Speakers of World Englishes remain drastically underrepresented, creating what Hovy and Spruit (2016) term 'exclusion by default' in natural language processing.

Linguistic imperialism in technology

Linguistic imperialism, as theorised by Phillipson (1992), refers to the dominance of one language over others, maintained through structural and ideological mechanisms that position the dominant language as inherently superior. Scholars have extended this framework to analyse how technologies reproduce linguistic hierarchies (Graham et al., 2011; Kornai, 2013). Technology-mediated linguistic imperialism operates through several mechanisms. Design decisions embed assumptions about standard language that privilege dominant varieties while marking others as deviant (Benjamin, 2019). The economic structure of technology development concentrates resources in wealthy Anglophone nations whose linguistic norms become universalised (Hecht & Gergle, 2010). Network effects create feedback loops where technologies optimised for dominant varieties attract more users from those groups, thereby generating more training data that further entrenches bias (Noble, 2018). Voice assistants exemplify these dynamics. By designing systems that comprehend dominant accents more accurately, technology companies create differential

access to digital services, effectively punishing speakers for using non-dominant varieties. Critical scholars argue that linguistic discrimination operates through what Flores and Rosa (2015) term 'raciolinguistic ideologies,' which position racialised speakers as inherently deficient language users regardless of their actual linguistic competence.

Research gap

Whereas existing research documents ASR bias against various accent groups, limited published studies have systematically examined voice assistant comprehension of Cameroonian English or other Central African varieties. Given Cameroon's unique bilingual context, its growing technology adoption, and the distinctive features of Cameroonian English, this variety provides a crucial test case for understanding how voice technologies perform with under-represented World Englishes. Moreover, previous studies have focused primarily on transcription accuracy without examining how phonological features specifically trigger recognition failures. This is the gap this research addresses through detailed error pattern analysis.

3. Methodology

This section outlines the mixed-methods research design employed to investigate voice assistant comprehension of Cameroonian English. The methodology integrates experimental phonetics to quantify accuracy disparities, acoustic analysis to identify phonological features associated with recognition failures, and qualitative interviews to understand user experiences of technological discrimination. This multifaceted approach enables both empirical measurement of comprehension gaps and critical examination of their sociolinguistic implications.

Research design

This study employed a mixed-methods approach combining experimental phonetics, quantitative error analysis, and qualitative user experience research. The experimental component measured voice assistant comprehension accuracy across three platforms, while error analysis identified phonological patterns in recognition failures. Qualitative interviews provided insight into user experiences and broader sociolinguistic implications.

Participants

Fifteen native speakers of Cameroonian English (8 female, 7 male; aged 22 to 48 years, M = 32.4, SD = 7.8) participated in the experimental phase. All participants were residents of Cameroon, recruited from Buea in the South West Region and Yaoundé, the capital city, through university networks, professional associations, and community contacts.

Table 1: Participant demographics

Characteristic		N	(%)
Gender	Female	8	53.3
	Male	7	46.7
Educational level	Secondary	3	20
	Undergraduate	6	40
	Postgraduate	6	40
Region	South West	9	60
	North West	4	26.7
	Other Anglophone	2	13.3
Primary Cameroonian Languages	Pidgin	15	100
	Mokpe	4	26.7
	Ewondo	3	20
	Other	8	53.3

Selection criteria included birth and primary education in Anglophone Cameroon, current regular use of English, no documented speech or hearing impairments, and at least occasional prior experience with voice assistants. Participants represented diverse educational backgrounds, ranging from secondary education to postgraduate degrees, and various socioeconomic statuses. Table 1 summarises participant demographics.

Five participants from the experimental group engaged in semi-structured interviews following experimental tasks to discuss their experiences with voice technologies.

Materials

The experimental protocol utilised three voice assistant platforms: Apple Siri (iOS 17.5), Google Assistant (Android 14.24), and Amazon Alexa (Echo Dot 5th generation). These platforms represent dominant voice assistant technologies globally and are increasingly available in Cameroonian markets.

The command set consisted of 50 standardised voice commands spanning five categories (10 commands each): basic queries ('What's the weather?', 'Set a timer for 10 minutes'), smart home controls ('Turn on the lights', 'Adjust temperature to 22 degrees'), information retrieval ('Define photosynthesis', 'Who wrote Things Fall Apart?'), navigation requests ('Directions to the nearest hospital in Tsinga'), and phonetically complex commands designed to test problematic phonological features ('Text test results', 'Tell us about Thomas Nkono').

Commands were selected to reflect realistic usage scenarios while incorporating phonological features characteristic of Cameroonian English that might trigger recognition errors. All commands were

validated for naturalness by three Cameroonian English speakers not involved in the study.

Procedure

Testing occurred in quiet environments with minimal background noise at research facilities in Buea and Yaoundé. Each participant completed the 50-command protocol with all three voice assistants, resulting in 150 commands per participant (2,250 total commands). Testing was conducted over three sessions to prevent fatigue, with each participant completing one platform per session and sessions separated by at least 24 hours. Command order was randomised within each session to control for order effects.

Commands were presented visually on cards, and participants read each command naturally. No instruction to modify accent or speaking style was given. Audio from each command was recorded using Shure SM58 microphones positioned 15 cm from speakers, allowing for subsequent acoustic analysis. Voice assistant responses were recorded and classified into four categories: correct comprehension (assistant executed intended command accurately), partial comprehension (assistant recognised some words but misunderstood command), miscomprehension (assistant executed incorrect command), and no recognition (assistant failed to register speech or reported inability to understand).

Two independent raters (both trained phoneticians) classified all responses. Inter-rater reliability was calculated using Cohen's kappa ($\kappa = 0.87$, 95% CI [0.84, 0.90]), indicating strong agreement. Discrepancies were resolved through discussion and consensus.

Phonological analysis

For commands resulting in comprehension failure, acoustic analysis was conducted using Praat phonetic analysis software version 6.3.16 (Boersma & Weenink, 2024). Analysis focused on four phonological parameters: rhythm metrics using Normalised Pairwise Variability Index (nPVI) to quantify syllable-timed versus stress-timed rhythm; consonant cluster realisation through spectrographic analysis of word-final clusters; vowel formants with F1 and F2 measurements for vowels in key lexical items; and prosodic contours examining fundamental frequency (F0) patterns for intonation analysis.

Measurements were compared against reference values from previous phonetic studies of Cameroonian English (Mbangwana, 2004; Simo Bobda, 1994) and contrasted with typical values for Standard American English to identify specific phonological features associated with recognition failures.

Interview protocol

Semi-structured interviews (M = 42 minutes, range = 35 to 58 minutes) explored participants' experiences with voice technologies. The interview guide included questions about prior experiences with voice assistants, emotional responses to comprehension failures, strategies employed when voice assistants fail to understand, perceptions of why comprehension failures occur, and broader reflections on language, technology, and identity.

Interviews were conducted in English, audio-recorded with permission, and transcribed verbatim. Thematic analysis followed Braun and Clarke's (2006) six-phase approach, with codes developed both deductively (based on theoretical frameworks of linguistic imperialism) and inductively (from patterns emerging in

participant narratives). Two assistants were recruited to independently code transcripts, achieving 89% agreement.

Statistical analysis

Comprehension accuracy data were analysed using one-way analysis of variance (ANOVA) to compare platforms, with post-hoc Tukey HSD tests for pairwise comparisons. Chi-square tests examined associations between phonological features and comprehension outcomes. Binary logistic regression assessed the relative contributions of phonological features to comprehension failure, with comprehension success (correct vs. failed) as the dependent variable and rhythm metrics, consonant cluster presence, vowel deviation scores, and prosodic deviation scores as predictors. Effect sizes are reported using Cohen's d for t-tests, η^2 for ANOVA, and odds ratios for logistic regression. Alpha was set at .05 for all tests.

Ethical considerations

All participants provided written informed consent and were compensated 10,000 CFA francs for their time. Participants were debriefed about the study's purpose and assured that comprehension failures reflected technological limitations rather than deficiencies in their speech. Given the potentially sensitive nature of discussions about linguistic discrimination, researchers emphasised that all English varieties are linguistically equal and that recognition failures constitute technological bias rather than speaker error.

4. Results

This section presents the empirical findings from the experimental and qualitative components of the study. The results are organized into four subsections: overall comprehension accuracy rates across

platforms, phonological features correlated with recognition failures, multivariate analysis of contributing factors, and qualitative findings from participant interviews.

Overall comprehension accuracy

Voice assistants demonstrated markedly lower comprehension rates for Cameroonian English compared to documented benchmarks for dominant varieties. Across all three platforms and 2,250 total commands, the overall correct comprehension rate was 56.8% (95% CI [54.9, 58.7]) with substantial variation between platforms (see Table 2).

Table 2: Comprehension accuracy by platform and error type

Platform	Correct (%)	Partial (%)	Miscomprehension (%)	No Recognition (%)
Google Assistant	63.4	14.2	13.1	9.3
Siri	54.7	16.8	17.2	11.3
Alexa	52.3	15.1	18.4	14.2
Overall	56.8	15.4	16.2	11.6

One-way ANOVA revealed significant differences between platforms, $F(2, 2247) = 87.34$, $p < .001$, $\eta^2 = .072$. Post-hoc Tukey HSD tests showed Google Assistant significantly outperformed both Siri ($p < .001$, $d = 0.45$) and Alexa ($p < .001$, $d = 0.58$). No significant difference was found between Siri and Alexa performance ($p = .187$, $d = 0.13$).

The analysis by command category revealed differential performance. Basic queries achieved highest accuracy (69.8%), followed by smart home controls (61.2%), information retrieval (53.7%), navigation (51.4%), and phonetically complex commands with especially words with local language sound clusters (31.6%).

The particularly low performance on phonetically complex commands suggests that specific phonological characteristics drive recognition errors.

These figures contrast sharply with documented accuracy rates for Standard American and British English, which consistently exceed 95% across similar command sets (Feng et al., 2021; Koenecke et al., 2020). The performance gap represents a 38.2 percentage point deficit (95% CI [36.1, 40.3]), translating to approximately 21.5 failed commands per 50-command protocol.

Phonological correlates of comprehension failure

Acoustic analysis of the 972 failed commands (43.2% of total attempts) identified four primary phonological features associated with recognition errors.

Syllable-timed rhythm

Cameroonian English exhibits syllable-timed rhythm, contrasting with the stress-timed rhythm of dominant varieties. Normalised Pairwise Variability Index (nPVI) values for participant speech averaged 46.7 (SD = 4.3), compared to typical Standard American English values of 65 to 70 (Low et al., 2000). Commands with greater syllable-timing showed significantly higher failure rates. Pearson correlation revealed a strong negative relationship between nPVI values and comprehension success, $r = -.68$, $p < .001$, meaning more syllable-timed utterances were less likely to be understood. For example, the command 'Set a timer for fifteen minutes' was frequently misrecognised. Cameroonian speakers produced relatively even stress across syllables (nPVI = 44.2), whilst ASR systems trained on stress-timed varieties expect prominent stress on timer, fifteen, and minutes with reduced vowels in unstressed positions.

Consonant cluster simplification

In the experimental data, 74.8% of word-final consonant clusters underwent some modification, including complete cluster reduction (for example, test pronounced without the final consonant), epenthesis with vowel insertion (as in film becoming three syllables), and devoicing of final consonants. Commands containing words with final clusters showed significantly higher failure rates (67.3%) compared to cluster-free commands (43.6%), $\chi^2(1) = 142.7$, $p < .001$, $\varphi = .25$. Some examples include: 'List the main points = [lis],' 'contrast the two theories' = [kontras, teories], 'extract' = [ekstrak]. The phonetically complex command 'Text test results,' which contains three word-final clusters, achieved only 13.3% comprehension accuracy. Spectrographic analysis confirmed systematic cluster modifications, yet ASR systems consistently failed to map these modified forms to intended words.

Vowel quality differences

Vowel formant analysis revealed systematic differences from Standard American English values. The diphthong /eɪ/ in words like 'play' and 'straight' was typically realised as a monophthong [e:] by 81.3% of participants, with formant measurements showing stable values throughout vowel duration (F1 = 420 Hz, F2 = 2180 Hz) rather than expected formant movement. Additionally, the vowel /æ/ in words like 'cat' showed significantly different formant patterns (F1 = 750 Hz, F2 = 1620 Hz vs. Standard American English F1 = 660 Hz, F2 = 1720 Hz), indicating a lower, more backed realisation.

Commands containing these vowels showed elevated error rates: 61.4% for words with /eɪ/ and 56.8% for words with /æ/, compared to 49.7% for commands with other vowels, $\chi^2(2) = 38.4$, $p < .001$,

Cramer's V = .13. The command 'Play straight eight' achieved only 10.7% accuracy, with common misrecognitions including 'Play Street eat' and 'Play straight heat.'

Intonation patterns

Prosodic analysis revealed differences in intonation contours, particularly in question formation. While this feature showed weaker association with comprehension failure than segmental features, it likely contributed to cumulative recognition difficulties when combined with other phonological differences.

Multivariate analysis

Binary logistic regression examined the relative contributions of phonological features to comprehension failure (see Table 3). The model significantly predicted comprehension outcomes, $\chi^2(4) = 386.2$, $p < .001$, Nagelkerke $R^2 = .34$, correctly classifying 71.2% of cases.

Table 3: Logistic regression predicting comprehension failure

Predictor	B	SE	Wald	p	OR	95%CI
Consonant cluster presence	0.89	0.12	54.3	<.001	2.44	[1.93,3.08]
nPVI (rhythm)	-0.06	0.01	36.7	<.001	0.94	[0.92,0.96]
Vowel deviation score	0.43	0.08	28.9	<.001	1.54	[1.32,1.80]
Prosodic deviation	0.11	0.54	51.2	.221	1.12	[0.94,1.33]
Constant	3.87	0.54	51.2.	<.001	48.05	--

Note. OR = odds ratio. Higher vowel and prosodic deviation scores indicate greater difference from Standard American English norms.

Significant individual predictors included consonant cluster presence (speakers producing clusters were 2.44 times more likely to experience comprehension failure), syllable-timing rhythm

(each unit decrease in nPVI increased failure odds by 6%), and vowel deviation scores (each unit increase in deviation increased failure odds by 54%). Prosodic deviation was not a significant independent predictor when other features were controlled. These findings confirm that segmental features and rhythmic timing represent the primary phonological obstacles to voice assistant comprehension of Cameroonian English.

User experience findings

Thematic analysis of interviews revealed four major themes regarding participant experiences with voice assistants.

Technological alienation and frustration

All five interview participants described frustration with voice technologies, using terms like 'irritating,' 'pointless,' and 'not made for us.' One participant (P3, 28-year-old female professional) stated:

> *'Every time I try Siri, it's like talking to someone who doesn't want to understand you. I can say something three times, clearly, and it still gives me nonsense. Eventually I just give up and type it. It makes you feel like your voice doesn't matter.'*

This sense of technological alienation extended beyond inconvenience to feelings of exclusion. Participants explicitly connected voice assistant failures to broader experiences of linguistic marginalisation.

Accent modification and linguistic shame

Four participants reported attempting to modify their accent to improve comprehension, describing this as 'speaking like an American' or 'doing the British thing.' Strategies included exaggerating stress contrasts, producing consonant clusters fully,

imitating different vowel sounds, and slowing speech rate significantly. One participant (P1, 34-year-old male) described:

> *'I find myself almost performing, you know? Like I'm doing an accent. I'm from Bamenda, I studied in England, I speak perfect English. But for my phone, I have to pretend to be someone else. It's exhausting, and honestly, it's embarrassing, generating laughter from bystanders even. Why should I have to change how I speak?'*

These accounts reveal how technological design choices impose psychological burdens on users, requiring them to perform linguistic assimilation to access services.

From self-blame to critical consciousness

Participants offered diverse explanations for why voice assistants fail to understand them. Three participants initially attributed failures to their own speech (*'Maybe I'm not speaking clearly enough'*). However, when prompted to reflect more deeply, all participants recognised the issue as technological bias. One participant (P5, 29-year-old female) stated:

> *'Actually, thinking about it now, the problem isn't me. I communicate perfectly fine with people. The problem is they didn't programme these things to understand different accents. They only care about American and British English.'*

This shift from self-blame to systemic critique suggests that initial failure attribution to oneself may reflect internalised linguistic hierarchies.

Digital exclusion and inequality

All participants connected voice assistant failures to broader patterns of technological inequality. Others noted that voice

interfaces are becoming mandatory for certain services. One participant (P4, 31-year-old female) observed:

> *'Everything is moving to voice control. But if voice control doesn't work for you, then what? You're locked out. It's like they're saying, this technology isn't for you.'*

Two participants described abandoning voice-activated devices they had purchased because of persistent comprehension failures, representing wasted financial investment.

5. Discussion

This section interprets the study's findings through sociolinguistic frameworks, and examines the broader implications of voice assistant comprehension disparities for Cameroonian English speakers. First, it presents an analysis of the disparities constituting sociolinguistic exclusion before exploring the role of specific phonological features in driving recognition failures. These patterns are subsequently framed as technological linguistic imperialism. The section concludes with implications for language education and technology policy.

Comprehension disparities as sociolinguistic exclusion

The finding that voice assistants comprehend only 56.8% of Cameroonian English commands, compared to over 95% accuracy for dominant varieties, represents a substantial and consequential disparity. This almost 38-point gap renders voice assistants largely unusable for Cameroonian English speakers. A success rate below 60% means nearly half of all commands fail, a user experience so unreliable that rational users would abandon the technology, which interview data confirmed.

This disparity reflects how technological systems encode prestige hierarchies. Standard language ideology, which positions certain

varieties as neutral and correct while treating others as marked and incorrect, becomes literally encoded into algorithmic systems. When ASR models are trained predominantly on Standard American English or British English data, they learn to recognise that variety as the norm against which all other input is evaluated. Cameroonian English, with its systematic phonological differences, is then treated not as a legitimate variety but as deviant input, or noise to be filtered out rather than a language to be understood.

The magnitude of the comprehension gap reveals what sociolinguists term linguistic discrimination, referring to the differential treatment based on language variety. Unlike overt discrimination where speakers are explicitly told their variety is unacceptable, algorithmic discrimination operates invisibly. The system simply fails to work, and without understanding how ASR functions, users may internalise this failure as personal inadequacy rather than recognising it as systemic bias.

Platform differences also merit interpretation. Google Assistant's superior performance (63.4%), whilst still inadequate, may reflect Google's larger and more globally diverse user base. However, even Google's best performance falls drastically short of usability standards, highlighting that incremental improvement within current training paradigms cannot solve this problem. Fundamental restructuring is required that centres on linguistic diversity.

Phonological features and recognition failure

This study's multivariate analysis identified consonant cluster simplification, syllable-timed rhythm, and vowel quality differences as primary predictors of comprehension failure. Understanding these features requires sociolinguistic perspective

on language variation. These phonological characteristics are not errors or deficiencies, rather they are systematic features of Cameroonian English arising from natural sociolinguistic processes (Kouega, 2007; Simo Bobda, 1994). Syllable-timing reflects substrate influence from Niger-Congo and Afro-Asiatic languages and consonant cluster simplification follows universal phonological principles favouring simpler syllable structures. Vowel quality differences reflect both substrate influence and adaptation to local phonological ecologies.

From a variationist sociolinguistic perspective, these features represent stable community norms, and shared patterns that mark membership in the Cameroonian English speech community. The fact that over 80% of participants demonstrated monophthongisation and over 74% simplified consonant clusters confirm these are systematic community features, not random variations. Yet ASR systems treat these systematic features as errors because they deviate from norms encoded in training data.

This reveals how technological systems participate in the social construction of linguistic legitimacy. By failing to recognise Cameroonian English features, voice assistants implicitly mark them as non-standard, thus reinforcing ideologies that devalue linguistic diversity. The strong correlations between specific features and comprehension failure demonstrate that recognition errors are not random but patterned, targeting features that diverge from dominant variety norms. This systematic bias parallels other forms of linguistic discrimination documented in sociolinguistic research (Rosa & Flores, 2017).

Technological linguistic imperialism

The systematic exclusion of Cameroonian English from effective voice interaction constitutes technological linguistic imperialism,

which refers to the reproduction and amplification of linguistic hierarchies through algorithmic systems. This extends Phillipson's (1992) framework to the digital realm, examining how technologies encode and enforce linguistic power relations through three interconnected mechanisms.

First, exclusion by default reflects how power shapes knowledge production. ASR development prioritises data from accessible speaker populations in technology hubs, creating training datasets that disproportionately represent dominant varieties (Blodgett et al., 2020). This structural bias produces discriminatory outcomes. From a political economy of language perspective, and through this exclusion reflects broader patterns where resources flow to already-privileged groups (Hecht & Gergle, 2010).

Secondly, differential access creates what Benjamin (2019) terms 'technical stratification,' an inequality embedded in technological infrastructure. Voice interfaces increasingly mediate access to essential services such as information, communication, smart home controls, accessibility features, customer service, and more. When these interfaces function poorly for certain accent groups, they create access barriers with material consequences. This differential access reinforces existing inequalities, with speakers of dominant varieties gaining seamless access while speakers of marginalised varieties face compounding barriers.

Lastly is the notion of identity taxation which refers to the psychological and cognitive burden imposed on speakers who must modify their linguistic production to accommodate technological limitations. Our interview data revealed that Cameroonian English speakers engage in accent modification to improve comprehension. However, this accommodation is not reciprocal. Speakers of dominant varieties need not adjust their

speech, whereas speakers of marginalised varieties must constantly monitor and modify their linguistic production.

From a language ideology perspective, this asymmetry reinforces perceptions that dominant varieties are natural and neutral whereas other varieties require adjustment. When speakers must perform linguistic assimilation to use technology, it sends powerful messages about whose language is valued, whose voice matters, and who belongs in digital spaces.

These mechanisms parallel historical linguistic imperialism where colonial powers positioned their languages as superior whilst devaluing indigenous languages and local varieties (Pennycook, 2017; Phillipson, 1992). Voice assistants represent a new frontier for these dynamics, now encoded into algorithms that govern access to digital resources.

Implications for language education and technology policy

These findings carry significant implications for language education. Voice assistants are increasingly positioned as language learning tools, yet our findings reveal that these tools systematically disadvantage learners with non-dominant accents, creating a double standard that reinforces linguistic hierarchies. For learners of English in Cameroon, voice assistants send implicit messages that their English is inadequate. When learners successfully communicate with human interlocutors but repeatedly fail with voice assistants, this teaches destructive lessons about linguistic legitimacy, potentially reinforcing the myth that only Inner Circle varieties constitute 'real' or 'correct' English (Canagarajah, 2013; Kachru, 1985).

Language educators must therefore approach voice assistants critically, recognising both their potential and profound limitations. Pedagogical responses should include teaching critical

digital literacy, and helping students analyse how technologies encode linguistic norms. Educators can also employ variety-affirming pedagogy, using voice assistant failures as opportunities to discuss linguistic diversity and legitimacy. If students wish to use voice assistants, educators can teach strategic code-switching that will aid in framing accent adjustment as a strategic skill rather than correction of errors, while maintaining that their natural variety is equally valid. Crucially, educators should resist positioning voice assistant comprehension as a measure of learner proficiency. Inability to be understood by Siri does not indicate inadequate English but rather inadequate technology.

Finally at policy levels, these findings suggest the need for regulatory frameworks ensuring linguistic accessibility in digital technologies. Just as physical accessibility standards require built environments to accommodate diverse bodies, linguistic accessibility standards could require voice technologies to accommodate diverse varieties, specifying minimum performance thresholds across accent groups and requiring transparency about training data composition.

6. Limitations and future research

With regard to the limitations of this study, first, the sample size (n = 15) was sufficient for experimental phonetics research but limits generalisability. Future studies should include larger samples across more diverse geographic regions, as the participants were primarily from Southwest Cameroon. Regional variation within Cameroonian English exists, with differences between Southwest and Northwest varieties, and these findings may not fully represent this diversity. Again, this study did not consider a control group of Standard American or British English speakers for direct comparison, relying instead on published

benchmarks. While these benchmarks are well-established in the literature, direct comparison under identical testing conditions would strengthen claims about differential performance.

Another limitation is established at the level where testing occurred in quiet environments that may not reflect real-world usage contexts. Voice assistants are often used in noisy environments (kitchens, cars, public spaces), where background noise may interact with accent features to further degrade performance. Future research should examine comprehension rates under varied acoustic conditions.

Again, this study was conducted at a single time point in 2024. Voice assistant technologies are continuously updated, and ASR models may improve over time. Longitudinal research tracking changes in comprehension accuracy as systems evolve would provide insight into whether technological progress naturally addresses linguistic bias or whether explicit intervention is required.

Finally, the study also focused exclusively on Cameroonian English. Comparative studies examining other African English varieties (Nigerian English, Kenyan English, South African Black English) would reveal whether the findings reflect broader patterns of ASR bias against African varieties or are specific to Cameroonian phonology. Similarly, research on other marginalised World Englishes (Caribbean English varieties, Southeast Asian varieties) would illuminate the global scope of this problem.

7. Conclusion

This research demonstrates that mainstream voice assistants exhibit severely degraded performance when comprehending

Cameroonian English, achieving only 56.8% accuracy compared to over 95% for dominant varieties. This disparity stems from specific phonological features, including syllable-timed rhythm, consonant cluster simplification, and distinctive vowel realisations that deviate from norms encoded in ASR training data. These features are systematic components of Cameroonian English, a legitimate variety with millions of speakers, yet systems treat them as errors.

Analysed through sociolinguistic frameworks, this study reveals how voice assistants function as mechanisms of technological linguistic imperialism (Phillipson, 1992). They create differential access to digital services, impose psychological burdens on speakers who must modify their accents, and reinforce ideologies positioning dominant varieties as inherently superior (Rosa & Flores, 2017). The invisibility of this process, mediated through algorithms rather than explicit policy, makes it particularly insidious, and thus obscuring the exercise of linguistic power.

The question is not whether machine learning can accommodate accent diversity. Technically, it can through diversified training data, pronunciation modelling, transfer learning, and adaptation features (Jain et al., 2018). The question is whether technology companies will prioritise linguistic inclusion, allocating resources to diversify training data and redesign systems around principles of linguistic justice rather than privileging easily accessible speaker populations in wealthy Anglophone nations. Consequently, for language education, these findings underscore the need for critical engagement with technology. Voice assistants should not be treated as neutral tools or arbiters of correct pronunciation. Instead, educators must help learners understand how technologies encode power relations, maintain that World

Englishes are legitimate varieties (Canagarajah, 2013; Kachru, 1985), and advocate for more inclusive technological design.

8. Recommendations

Based on the findings, the study recommends that the technological companies diversify training data to include substantial representation of World Englishes varieties, with particular attention to underrepresented African varieties. Companies should implement regular bias audits measuring performance across accent groups and making results publicly available. They should establish minimum performance thresholds requiring that ASR systems achieve at least 90% accuracy across all major English varieties before deployment. Investment in accent-agnostic ASR architectures that can adapt to speaker characteristics without requiring massive variety-specific datasets is essential. Finally, companies should engage in community-based participatory design with speakers of marginalised varieties to understand their needs and priorities.

For policymakers, the study recommends establishing linguistic accessibility standards for commercial voice technologies, modelled on existing accessibility regulations for physical and digital environments. Governments should require transparency in ASR training data composition, mandating disclosure of which varieties are represented and in what proportions. Funding for research on inclusive ASR technologies should be prioritised, particularly research conducted by scholars from underrepresented linguistic communities. Educational technology procurement policies should specify linguistic accessibility requirements, ensuring that schools do not adopt technologies that discriminate against local varieties.

For educators, the study recommends teaching critical digital literacy that helps students understand how algorithmic systems encode social biases. Educators should explicitly affirm the legitimacy of World Englishes in curriculum and assessment, resisting pressures to position voice assistant comprehension as a measure of proficiency. They should advocate for institutional technology policies that consider linguistic diversity. Professional development should address linguistic ideologies and their reproduction through educational technologies.

For researchers, the study recommends expanding documentation of ASR bias across World Englishes varieties, building an evidence base that can inform policy and technical interventions. Collaboration with computer scientists to develop and test bias reduction techniques is essential. Researchers should conduct more extensive qualitative research on user experiences of technological linguistic discrimination. Critical analysis of the political economy of ASR development, examining how market structures and corporate strategies perpetuate linguistic hierarchies, can illuminate paths toward change.

As digital interfaces become increasingly voice-mediated, the stakes of this exclusion escalate. Access to information, services, and opportunities cannot be predicated on speaking a particular variety of English. The technological capacity to build inclusive systems exists. What remains is the political will to demand and implement linguistic justice in artificial intelligence. This study joins growing scholarship documenting how algorithmic systems reproduce social hierarchies (Benjamin, 2019; Blodgett et al., 2020; Noble, 2018). The challenge before researchers, educators, policymakers, and technology developers is to insist that linguistic diversity is not a technical inconvenience to be minimised but a

fundamental dimension of human communication to be valued and accommodated. Only through such commitment can we build technological futures that serve all speakers equitably, regardless of their accent.

References

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5454-5476.

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version 6.3.16) [Computer software]. http://www.praat.org/

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3(2)*, 77-101.

Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations.* Routledge.

Feng, S., Kudina, O., Halpern, B. M., & Scharenborg, O. (2021). *Quantifying bias in automatic speech recognition*. arXiv preprint arXiv:2103.15122.

Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review, 85(2),* 149-171.

Graham, M., Hale, S. A., & Stephens, M. (2011). *Geographies of the world's knowledge*. Convoco! Edition.

Hecht, B., & Gergle, D. (2010). The tower of Babel meets web 2.0: User-generated content and its applications in a multilingual context. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 291-300.

Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,* 591-598.

Jain, A., Upreti, M., & Jyothi, P. (2018). Improved accented speech recognition using accent embeddings and multi-task learning. *Proceedings of Interspeech* 2018, 2454-2458.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge University Press.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences, 117(14),* 7684-7689.

Kornai, A. (2013). Digital language death. *PLOS ONE, 8(10)*, e77056.

Kouega, J. P. (2007). The language situation in Cameroon. *Current Issues in Language Planning, 8(1),* 3-94.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43(4),* 377-401.

Mbangwana, P. N. (2004). *English patterns of usage and usages in Cameroon*. Peter Lang.

Ngefac, A. (2008). *Social differentiation in Cameroon English: Evidence from sociolinguistic surveys*. Peter Lang.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Pennycook, A. (2017). *The cultural politics of English as an international language*. Routledge.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford University Press.

Rosa, J., & Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society, 46(5),* 621-647.

Simo Bobda, A. (1994). *Aspects of Cameroon English phonology*. Peter Lang.

Statista. (2024). Number of digital voice assistants in use worldwide from 2019 to 2024. https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/

Tatman, R. (2017). Gender and dialect bias in YouTube's automatic captions. *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, 53-59.