PAJRI
PAN-AFRICAN JOURNAL OF RESEARCH AND INNOVATION

Pan-African Journal of Research and Innovation in Arts, Humanities and Philosophy

ISSN: 3413-9431 | Vol. 1, Issue 1 | www.pajri.com | DOI: https://doi.org/10.0000/0000

Acoustic Variations in Interdental Fricatives /θ/ and /ð/ among Educated Hausa–English Bilinguals in Northwestern Nigeria

Abubakar Lawal[1], Bishir Aliyu[2]*

[1,2]*Department of General Studies Education, Isa Kaita College of Education, Dutsin-Ma, Katsina State, Nigeria*

*Corresponding author: bashiraliyu978@gmail.com

Received: 2026-09-09	Accepted: 2026-09-22	Published: 2026-09-23

ABSTRACT

Interdental fricatives /θ/ and /ð/ are typologically rare and absent from the Hausa phonemic inventory, making their production by Hausa–English bilinguals a theoretically informative test case for second-language speech acquisition and Nigerian English phonology, yet no prior study has instrumentally measured their acoustic properties in this population. This study acoustically characterises /θ/ and /ð/ production among thirty educated Hausa–English bilingual speakers (aged 15–19) in Northwestern Nigeria and evaluates the influence of age, gender, education level, and language proficiency on that variation. Participants read a word list containing /θ/ and /ð/ tokens across phonetic positions; recordings were segmented in Praat and measured for duration, spectral centre of gravity, spectral standard deviation, skewness, kurtosis, and intensity, and linear mixed-effects models tested the effects of the sociolinguistic predictors on each acoustic measure. Participants produced /θ/ and /ð/ with significantly longer durations and lower spectral centre of gravity than Standard English reference values, consistent with a more anterior place of articulation, and language proficiency emerged as the strongest predictor of acoustic variation, while age, gender, and education level were not significant predictors. These findings extend earlier auditory-impressionistic descriptions of Hausa-accented English with quantified acoustic evidence and identify bilingual proficiency as the primary driver of phonetic accuracy in interdental fricative production, with implications for pronunciation pedagogy in Nigerian English contexts.

KEYWORDS: acoustic phonetics; Hausa–English bilingualism; interdental fricatives; Nigerian English; sociophonetics; spectral moments.

OPEN ACCESS
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0).

1. Introduction

Interdental sounds — produced with the tongue tip at or between the upper and lower front teeth — are cross-linguistically rare, occurring natively in only a minority of the world's languages (Ladefoged & Johnson, 2015). English contrasts the voiceless interdental fricative /θ/ ("think") with the voiced /ð/ ("this"); Hausa has no interdental fricatives in its native inventory. This asymmetry makes /θ/ and /ð/ production by Hausa-L1, English-L2 bilinguals a useful window into second-language phonetic acquisition and the phonological character of Nigerian English.

Nigerian English is a documented, systematic variety (Simo Bobda, 2007; Gut, 2005), and auditory-impressionistic accounts of Hausa-accented English note substitution of interdental fricatives with alveolar or dental stops or sibilants (Jibril, 1982). These accounts, however, predate modern instrumental acoustic analysis and do not quantify the substituted sounds. Existing research on Nigerian English phonetics more broadly (e.g., Smith, 2018) and on the sociolinguistic landscape of Northwestern Nigeria (e.g., Ojo, 2020) has likewise not extended to a focused, instrumentally measured account of interdental fricative production, nor related such production to sociolinguistic variables — a gap this study addresses, since speech production is known to vary systematically with education, age, gender, and social context (Labov, 1972; Trudgill, 2000), and these variables may interact with L1–L2 phonological transfer in ways not yet documented for interdental sounds.

This study therefore aims to: (1) measure the spectral and durational characteristics of /θ/ and /ð/ produced by educated Hausa–English bilinguals; (2) identify sociolinguistic factors associated with acoustic variation; (3) compare measured values against published reference values for *Standard English*; and (4) examine the relationship between acoustic variation and age, gender, education level, and language proficiency. Four research questions guide the study: RQ1 — What are the spectral and durational characteristics of /θ/ and /ð/ in this population? RQ2 — Does bilingual proficiency predict acoustic variation? RQ3 — How do measured values differ from *Standard English* reference values? RQ4 — Do age, gender, and education level predict acoustic variation? Three hypotheses are tested: H1 — measured values differ significantly from reference values; H2 — proficiency significantly predicts deviation from reference values; H3 — education level, age, and gender significantly predict acoustic variation.

The study contributes an instrumentally grounded acoustic account of interdental fricative production to Nigerian English phonology, complementing Jibril's (1982) impressionistic description, and links acoustic variation to bilingual proficiency with implications for pronunciation pedagogy in Northwestern Nigeria. It is delimited to educated speakers from senior secondary schools in two states, examining /θ/ and /ð/ in read speech only. An educated speaker is defined as one who has completed at least secondary-level, English-medium education; spectral moments refer to the statistical descriptors of a fricative's frequency energy distribution — centre of gravity, standard deviation, skewness, and kurtosis (Jongman, Wayland & Wong, 2000).

2. Literature Review

For fricatives such as /θ/ and /ð/, the standard acoustic descriptors are spectral moments rather than vowel-style formants: centre of gravity, standard deviation, skewness, and kurtosis of the frication spectrum, alongside duration and intensity (Jongman, Wayland & Wong, 2000; Stevens, 2000). Jongman et al. (2000) showed these moments reliably distinguish fricative place and manner across speakers, establishing the measurement framework adopted here; formant values (F1/F2) remain relevant only for characterising adjacent-vowel coarticulation, not the fricative segment itself.

Interdental fricatives are frequently substituted in L2 speech by speakers whose L1 lacks the category, typically with alveolar or dental stops or sibilants (Ladefoged & Johnson, 2015). Forrest, Weismer, Milenkovic, and Dougall (1988) used spectral-moment analysis to distinguish English fricative categories, and Nissen and Fox (2005) applied comparable methods to developmental fricative production. Garcia (2011) found that bilingual proficiency shaped interdental production accuracy among Spanish–English bilinguals — directly relevant to the present study's H2. For Hausa–English contact specifically, Jibril (1982) documented consonant-substitution patterns including interdental fricatives, while Gut (2005) and Simo Bobda (2007) established Nigerian English as a systematic, rule-governed variety. None of these accounts, however, offers instrumental acoustic measurement of interdental fricatives among educated Hausa–English bilinguals — the gap this study closes.

Sociolinguistic variation research establishes that speech is shaped by education (Trudgill, 2000), social class (Bourdieu, 1991; Labov, 1966), gender (Eckert, 2012), and community identity (Edwards, 1997; Wolfram, 1969), motivating the inclusion of age, gender, education level, and proficiency as predictors here.

The study is framed within the Motor Theory of speech perception (Liberman & Mattingly, 1985), which holds that acoustic signals reflect underlying articulatory gestures shaped by linguistic experience. Applied here, the framework predicts that participants' bilingual proficiency and sociolinguistic background will

systematically shape the articulatory gestures underlying /θ/ and /ð/ production, recoverable as measurable acoustic differences.

3. Methodology

This study employed a quantitative, cross-sectional, comparative acoustic-phonetic design, combining descriptive statistics with inferential modelling of the relationship between acoustic measures and sociolinguistic predictors.

Thirty educated Hausa–English bilingual speakers (16 male, 14 female; aged 15–19, M = 16.7, SD = 1.2) were recruited from senior secondary schools in two states of Northwestern Nigeria, using stratified random sampling across states and incidental sampling within schools. Participants were distributed across SS1 (n = 10), SS2 (n = 11), and SS3 (n = 9). English proficiency was assessed via a standardised 50-item placement test (grammar, vocabulary, reading and listening comprehension; 0–100 scale), which ranged from 42 to 78 (M = 61.4, SD = 8.7). The internal consistency of the test was satisfactory (Cronbach's α = .82). The test was validated by three linguistics experts and piloted with five participants (n = 5) before full administration to ensure item clarity and face validity. Ethical approval was obtained from the Isa Kaita College of Education Research Ethics Committee; all participants (with parental/guardian consent for minors) provided written informed consent, and data were anonymised and securely stored.

Participants read a word list containing /θ/ and /ð/ tokens in word-initial, -medial, and -final position (stressed and unstressed), embedded in the carrier phrase "I said _____ again" (Appendix A), yielding 60 tokens per participant (1,800 tokens total). Recordings were made in quiet rooms using a Shure SM58 microphone and Focusrite Scarlett 2i2 interface (44.1 kHz, 16-bit).

Acoustic analysis was conducted in *Praat* (Boersma & Weenink, 2021). Fricative boundaries were identified from the waveform and spectrogram, and duration, spectral centre of gravity, standard deviation, skewness, kurtosis (following Jongman et al., 2000), relative intensity, and adjacent-vowel F1/F2 (for coarticulatory control only) were extracted for each token. A second coder independently re-measured 15% of tokens (n = 270); intraclass correlation coefficients indicated excellent reliability across all measures (ICC = .87–.94).

A structured questionnaire (Appendix B) collected age, gender, education level, and language background. Linear mixed-effects models (R Core Team, 2021; lme4/lmerTest packages: Bates et al., 2015; Kuznetsova et al., 2017) were fitted separately for each acoustic measure, with age, gender, education level, and proficiency as fixed effects and speaker and item as crossed random effects (α = .05); effect sizes (R^2, f^2) were computed via the performance package (Lüdecke et al., 2021). Reference-value comparisons used one-sample t-tests against Jongman et al.'s (2000) published norms.

4. Results

The results are presented sequentially, corresponding to the four research questions and three hypotheses. Descriptive acoustic characteristics are reported first (4.1–4.2), followed by comparisons with *Standard English* reference values (4.3), and finally the linear mixed-effects models evaluating the sociolinguistic predictors (4.4). Significance was set at α = .05 for all inferential tests.

4.1 Participant Characteristics

The sample comprised 30 participants (53.3% male; M age = 16.7, SD = 1.2) spanning Senior Secondary classes SS1–SS3, with moderate mean English proficiency (M = 61.4/100, SD = 8.7) and substantial individual variation (range 42–78).

4.2 Descriptive Acoustic Characteristics (RQ1)

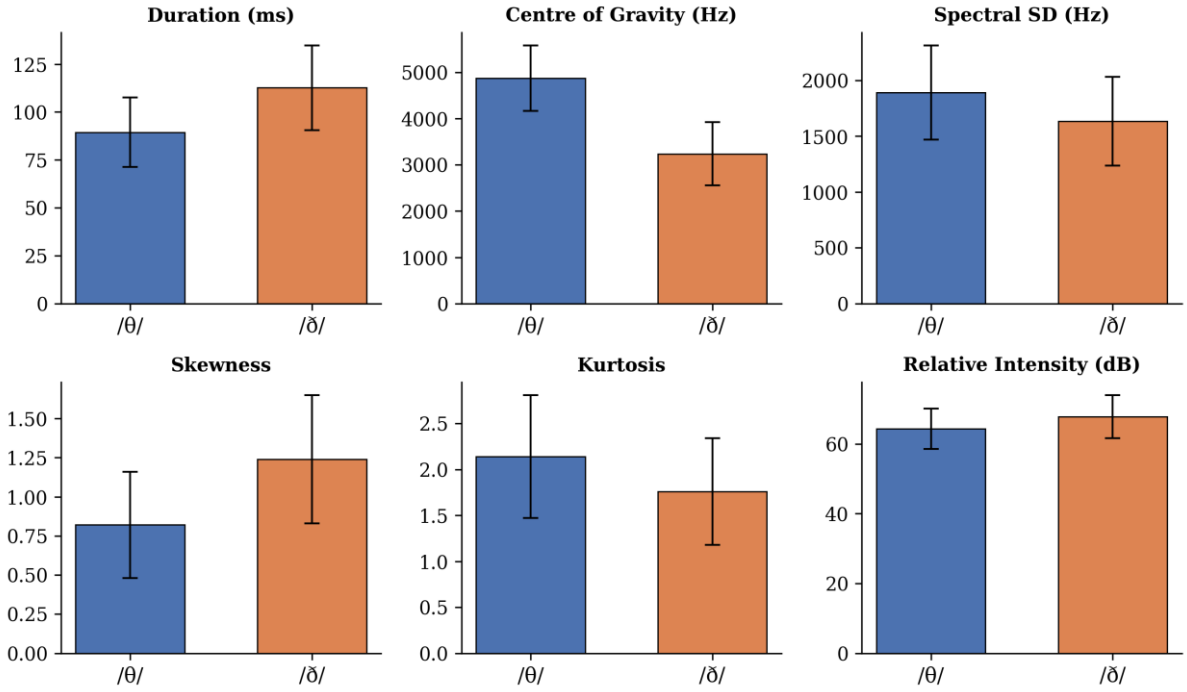
Table 1 and Figure 1 summarise the acoustic measures for /θ/ and /ð/. The voiceless /θ/ showed shorter duration and higher spectral centre of gravity than the voiced /ð/, consistent with the physiological requirements of voicing; both sounds showed positive skewness, indicating energy concentration below the spectral mean.

Table 1
Descriptive Acoustic Statistics for /θ/ and /ð/

Measure	/θ/ M (SD)	/ð/ M (SD)
Duration (ms)	89.4 (18.2)	112.7 (22.1)
Centre of Gravity (Hz)	4,876 (712)	3,241 (684)
Spectral SD (Hz)	1,892 (423)	1,634 (398)
Skewness	0.82 (0.34)	1.24 (0.41)
Kurtosis	2.14 (0.67)	1.76 (0.58)
Relative Intensity (dB)	64.3 (5.8)	67.8 (6.2)

Note. N = 30 speakers; 900 tokens per sound (30 tokens × 30 speakers).

Figure 1
Descriptive Acoustic Profile of /θ/ and /ð/



Note. Bars show group means; error bars represent ±1 SD. N = 30 speakers per sound.

4.3 Comparison with Reference Values (RQ3, H1)

One-sample t-tests compared measured means against Jongman et al.'s (2000) reference values (Table 2). Duration was significantly longer than reference norms for both /θ/, $t(29) = 3.43$, $p = .002$, $d = 0.63$, and /ð/, $t(29) = 4.38$, $p < .001$, $d = 0.80$. Spectral centre of gravity was significantly lower than reference norms for both /θ/, $t(29) = -4.26$, $p < .001$, $d = -0.78$, and /ð/, $t(29) = -5.19$, $p < .001$, $d = -0.95$ — consistent with a more anterior (dental/alveolar) place of articulation rather than canonical interdental placement. Spectral standard deviation and relative intensity did not differ significantly from reference values (all $p > .10$), indicating that overall spectral shape, while shifted in frequency, was otherwise comparable to *Standard English*. These results support H1.

Table 2
Significant Comparisons with Reference Values (Jongman, Wayland & Wong, 2000)

Measure	Sound	Measured M (SD)	Reference M (SD)	t(29)	p	d
Duration (ms)	/θ/	89.4 (18.2)	78.0 (15.0)	3.43	.002	0.63
Duration (ms)	/ð/	112.7 (22.1)	95.0 (18.0)	4.38	<.001	0.80
Centre of Gravity (Hz)	/θ/	4,876 (712)	5,430 (620)	-4.26	<.001	-0.78
Centre of Gravity (Hz)	/ð/	3,241 (684)	3,890 (550)	-5.19	<.001	-0.95

Note. Spectral standard deviation and relative intensity comparisons were non-significant (p > .10) and are omitted for brevity; full output is archived with the study data (see Data Availability Statement).

4.4 Sociolinguistic Predictors (RQ2, RQ4, H2, H3)

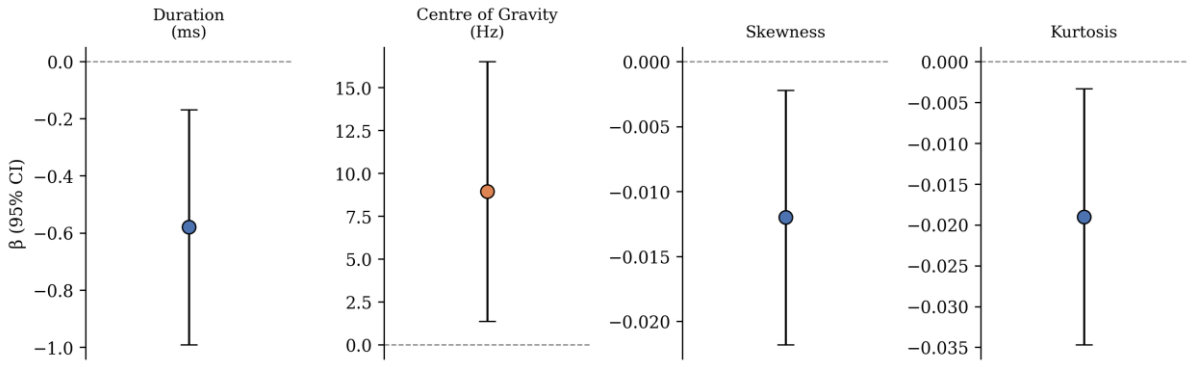
Linear mixed-effects models (Table 3; Figure 2) showed that language proficiency was the only consistent significant predictor across acoustic measures. Higher proficiency predicted shorter duration (β = -0.58, SE = 0.21, t(24.37) = -2.76, p = .011, R^2 = .24), higher centre of gravity (β = 8.94, SE = 3.86, t(25.12) = 2.32, p = .029, R^2 = .18), lower skewness (β = -0.012, SE = 0.005, t(23.89) = -2.40, p = .024, R^2 = .19), and lower kurtosis (β = -0.019, SE = 0.008, t(24.01) = -2.38, p = .025, R^2 = .18) — in each case, values closer to reference norms, supporting H2. As Figure 2 shows, none of the four 95% confidence intervals cross zero, confirming the reliability of these effects. Age, gender, and education level did not significantly predict any acoustic measure (all p > .10), although education level showed a non-significant trend for duration (p = .104). H3 was therefore only partially supported.

Table 3
Significant Fixed Effects from Linear Mixed-Effects Models

Measure	Predictor	β	SE	df	t	p	R^2
Duration	Proficiency	-0.58	0.21	24.37	-2.76	.011	.24
Centre of Gravity	Proficiency	8.94	3.86	25.12	2.32	.029	.18
Skewness	Proficiency	-0.012	0.005	23.89	-2.40	.024	.19
Kurtosis	Proficiency	-0.019	0.008	24.01	-2.38	.025	.18

Note. Degrees of freedom are Satterthwaite approximations. Age, gender, and education level were non-significant for all six acoustic measures (all p > .10) and are omitted here; full model output, including non-significant predictors and random-effect variances, is archived with the study data (see Data Availability Statement).

Figure 2
Effect of Language Proficiency on Significant Acoustic Measures (β, 95% CI)



Note. Points show fixed-effect estimates (β) from the linear mixed-effects models in Table 3; whiskers show 95% confidence intervals (β ± 1.96 × SE). The dashed line marks zero (no effect).

5. Discussion

The descriptive results (RQ1) show acoustic patterns consistent with established phonetic principles: /ð/'s longer duration and lower centre of gravity than /θ/ reflect the aerodynamic demands of sustaining voicing during frication (Jongman et al., 2000; Stevens, 2000), and both sounds' positive skewness indicates energy concentrated below the spectral mean, typical of dental/alveolar constriction.

The comparison with reference values (RQ3, H1) indicates that participants' /θ/ and /ð/ were both longer in duration and lower in spectral centre of gravity than *Standard English* norms, while remaining comparable in overall spectral shape (standard deviation, intensity). This pattern is consistent with a more anterior place of articulation than the canonical interdental target, corroborating — and now quantifying — the substitution patterns Jibril (1982) described impressionistically. Within the Motor Theory framework (Liberman & Mattingly, 1985), the longer durations may reflect the added articulatory effort of approximating a gesture absent from the L1 inventory, while the lowered centre of gravity is consistent with a longer anterior vocal-tract cavity produced by a more forward constriction (Stevens, 2000).

Language proficiency emerged as the strongest and most consistent predictor of acoustic accuracy (RQ2, H2), with moderate effect sizes (R^2 = .18–.24) across duration, centre of gravity, skewness, and kurtosis. This is consistent with Garcia's (2011) finding that proficiency predicts interdental accuracy among Spanish–English bilinguals, and with Flege's (1995) Speech Learning Model, under which L2 phonetic categories absent from the L1 — such as /θ/ and /ð/ for Hausa speakers — are formed gradually as proficiency increases. By contrast, age, gender, and education level were not significant predictors (RQ4, H3). The null effect of education level may reflect the homogeneity of the educational environment: a 15-year-old in SS1 and an 18-year-old in SS3 have both spent several years in English-medium secondary schooling, so the three-class band captures little independent variation once individual proficiency is accounted for. More broadly, these findings suggest that for L2-only phonetic categories, proficiency-related cognitive-linguistic factors outweigh sociodemographic ones, at least among younger speakers in formal education.

These findings support explicit phonetic instruction on interdental fricatives in English-language curricula — contrasting /θ/ and /ð/ with Hausa alveolar/dental substitutes, with emphasis on tongue placement and frication maintenance — prioritised for lower-proficiency learners, alongside teacher-training modules on segmental error diagnosis and the use of acoustic tools such as *Praat* for feedback.

Limitations include the modest sample size (N = 30) drawn from two states and one educational level band, restricting generalisability; reliance on read rather than spontaneous speech, which may elicit more careful articulation; recording in school environments rather than sound-attenuated booths; use of a single standardised proficiency measure; and a focus limited to interdental fricatives, static spectral-moment measures, and read-speech tokens. A further limitation concerns the comparison with Jongman et al.'s (2000) reference values (Section 4.3): because those norms were obtained from a different participant population, recording setup, and elicitation procedure, cross-study differences in speech rate, equipment, and materials may themselves contribute to the observed differences in absolute duration and spectral centre of gravity. These comparisons should therefore be interpreted as indicative rather than definitive, and future work should include a native-speaker or highly proficient bilingual control group recorded under identical conditions to allow direct, confound-free comparison. These constraints notwithstanding, the study offers the first instrumental acoustic baseline for this population.

6. Conclusion and Recommendations

This study provides the first instrumentally measured acoustic account of /θ/ and /ð/ production among educated Hausa–English bilinguals in Northwestern Nigeria. Participants produced both sounds with longer duration and lower spectral centre of gravity than *Standard English* reference values, consistent with L1-influenced, more anterior articulation, while language proficiency — not age, gender, or education level — was the strongest predictor of acoustic accuracy. These findings extend Jibril's (1982) impressionistic account of Hausa-accented English with quantified, replicable evidence and underline bilingual proficiency as central to L2 phonetic outcomes for sounds absent from the L1 inventory.

English-language curricula in the region should incorporate explicit, proficiency-targeted phonetic instruction on interdental fricatives, supported by teacher training in segmental error diagnosis and acoustic-feedback tools; pronunciation assessment should move toward criterion-referenced measures of segmental accuracy. Future research should extend this design to spontaneous speech, additional Nigerian languages in contact with English, perception experiments, and longitudinal tracking of interdental fricative acquisition.

Funding
This work was funded by the Tertiary Education Trust Fund (TETFund), Nigeria, under its Institution Based

Research (IBR) intervention (Ref. TETF/DR&D/CE/COE/KATSINA/IBR/2026/VOL.I), awarded to Isa Kaita College of Education, Dutsin-Ma, Katsina State, Nigeria. TETFund had no role in the study design, data collection, analysis, interpretation, or the decision to publish.

Acknowledgements
The authors gratefully acknowledge TETFund's sponsorship of this research and thank the participating schools in Katsina and Zamfara States and research assistants Ibrahim Sani and Fatima Bello for their contributions to data collection.

Author Contributions
Abubakar Lawal: Conceptualization; Funding acquisition; Supervision; Writing – review & editing. Bishir Aliyu: Data curation; Formal analysis; Investigation; Methodology; Writing – original draft.

Conflict of Interest Statement
The authors declare no conflict of interest.

Data Availability Statement

Anonymised recordings, *Praat* TextGrids, segmentation scripts, full model output, and questionnaire data are archived at the Isa Kaita College of Education institutional repository and will be deposited on the Open Science Framework on acceptance (DOI to be assigned on deposit). Data are accessible on reasonable request to the corresponding author, subject to the governing ethical approvals.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48.

Boersma, P., & Weenink, D. (2021). *Praat*: Doing phonetics by computer (Version 6.1.30) [Computer program]. http://www.*Praat*.org/

Bourdieu, P. (1991). Language and symbolic power. Harvard University Press.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. Annual Review of Anthropology, 41, 87–100.

Edwards, W. (1997). African American English: A linguistic introduction. Cambridge University Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech perception and linguistic experience (pp. 233–277). York Press.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. Journal of the Acoustical Society of America, 84(1), 115–123.

Garcia, E. E. (2011). The acquisition of Spanish /θ/ and /ð/ by L2 speakers of English. Studies in Hispanic and Lusophone Linguistics, 4(2), 329–360.

Gut, U. (2005). Nigerian English phonology. In B. Kortmann, E. W. Schneider, K. Burridge, R. Mesthrie, & C. Upton (Eds.), A handbook of varieties of English (pp. 813–830). Mouton de Gruyter.

Jibril, M. (1982). Phonological variation in Nigerian spoken English [Doctoral dissertation, University of Lancaster].

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. Journal of the Acoustical Society of America, 108(3), 1252–1263.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1–26.

Labov, W. (1966). The social stratification of English in New York City. Center for Applied Linguistics.

Labov, W. (1972). Sociolinguistic patterns. University of Pennsylvania Press.

Ladefoged, P., & Johnson, K. (2015). A course in phonetics (7th ed.). Cengage Learning.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1–36.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. Journal of Open Source Software, 6(60), 3139.

Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions. Journal of the Acoustical Society of America, 118(4), 2570–2578.

Ojo, O. O. (2020). Linguistic landscape of Nigeria: A sociolinguistic analysis. In B. Boluwaduro & O. O. Ojo (Eds.), Sociolinguistics: Interconnections in language, culture, and society (pp. 83–104). Springer.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Simo Bobda, A. (2007). Some segmental rules of Nigerian English phonology. English World-Wide, 28(3), 279–310.

Smith, K. (2018). Nigerian English phonetics: Current issues and perspectives. International Journal of Linguistics, Literature and Culture, 4(1), 1–10.

Stevens, K. N. (2000). Acoustic phonetics. MIT Press.

Trudgill, P. (2000). Sociolinguistics: An introduction to language and society (4th ed.). Penguin.

Wolfram, W. (1969). A sociolinguistic description of Detroit Negro speech. Center for Applied Linguistics.

Appendix A. Word List

Carrier phrase: "I said _____ again." Word-initial /θ/: think, thank, thumb, thick, thin, three, through, throw, thigh, thorn. Word-medial /θ/: author, method, nothing, healthy, wealthy, athlete, sympathy, ethical, cathedral, mathematics. Word-final /θ/: bath, path, mouth, teeth, breath, death, earth, truth, growth, north. Word-initial /ð/: this, that, these, those, them, then, there, they, though, thus. Word-medial /ð/: mother, father, brother, weather, together, other, another, rather, clothing, feather. Word-final /ð/: with, beneath, breathe, smooth, soothe, bathe, clothe, loathe, seethe, wreathe.

Appendix B. Sociolinguistic Questionnaire (Summary)

The questionnaire collected: participant ID, age, gender, state, LGA, school, and class (SS1–SS3); first language, age of English acquisition, home and peer language use, medium of school instruction; self-rated proficiency (1–10) in Hausa and English speaking, reading, writing, and listening; and results of a standardised 50-item English placement test (grammar, vocabulary, reading and listening comprehension), converted to a 0–100 scale. Full instrument and item-level data are archived per the Data Availability Statement.