Proceedings Paper:
HELLMUTH, SAM and Passoni, Elisa (2026) Multicultural London English Intonation: Pattern and Performance. In: Proceedings, Speech Prosody 2026. Speech Prosody, 26-29 May 2026 Speech Prosody. ISCA-INST SPEECH COMMUNICATION ASSOC, USA, pp. 407-411.

https://doi.org/10.21437/SpeechProsody.2026-82

Multicultural London English Intonation: Pattern and Performance

Sam Hellmuth[1], *Elisa Passoni*[2]

[1]University of York
[2]Queen Mary University of London

sam.hellmuth@york.ac.uk, elisa.passoni@qmul.ac.uk

Abstract

Multicultural London English (MLE) intonation awaits detailed description though its distinctiveness on other linguistic levels is well-established. MLE descriptions typically derive from sociolinguistic interviews, ill-suited to elaboration of an intonational description of a new variety. We propose a method for forming generalizations about an intonational system from (semi-)spontaneous interview data, where tokens for prosodic analysis are identified using non-prosodic criteria, to avoid circularity. We explore two features which each demonstrate a different type of selection criteria, applied to corpus data from eight MLE speakers from East London: i) polar questions, identified by 'next-turn-proof'; and ii) emphatic lengthening, identified from segmental phonetic cues. Identified tokens were auditorily annotated for contour shape (fall, rise, rise-fall, etc.) and segmental landmarks manually labelled. We use landmark registered fPCA to explore the mapping of F0 contours to auditory labels, controlling for stress position in target words.

The range of MLE polar question contours resembles that of Mainstream English Varieties but their relative frequency is highly distinct. Emphatic consonant lengthening (ECL) favors word-initial obstruent targets but the near-ubiquitous rise-fall contours on ECL-lengthened words and on polar questions are phonetically very similar. We note relevant features of substrate varieties and suggest features amenable to the same method.

Index Terms: intonational phonology, Multicultural London English, sociophonetic variation, discourse intonation

1. Introduction

1.1. Multicultural London English

About 30 years ago, MLE, a new linguistic variety of British English was documented in the context of highly diverse inner-city London neighborhoods, in the traditionally working-class East End of the city. MLE has been described as a multiethnolect born out of indirect language contact among ethnically-diverse adolescent friendship groups that "contains a core of innovative phonetic, grammatical and discourse-pragmatic features" [1, p. 65].

Segmental properties of MLE have been extensively documented, with key phonetic features ranging from unshifted, monophthongization of diphthongs, to TH- and DH-stopping (e.g. 'thing' [tɪŋg], 'them' [dɛm]), from a reduction in H-dropping to k-backing before non-high back vowels [2, 3]. More recent contemporary work has shown MLE is highly stable in its segmental phonological profile, however, and, differently from what was suggested in the earlier work, MLE carries clear, albeit subtle, phonetic variation arising from community-specific ethnic differences [4, 5].

1.2. Intonation of MLE and related varieties

Anecdotal reports describe an 'MLE intonation' characterized by wide and/or raised pitch range and sudden pitch changes [6], as well as the use of emphatic vowel lengthening [7], parallel emphatic consonantal lengthening [p.c. Paul Kerswill], and, in our own observations, a preference for declarative questions. Despite these converging reports of key features of interest, MLE intonation has to date received little attention.

The only available work on the intonation of London English is the analysis (and re-analysis) of a small dataset of inner-city London English as spoken by monolingual English teens of Caribbean heritage at the end of the 1990s (thus arguably MLE) from the IViE corpus [8-10]. Grabe's briefly reported generalizations in [8] implied that speakers of London English used two nuclear contours (rise L*_H%, high-plateau H*_H%), in different types of questions, which were not used by speakers of Southern Standard British English (SSBE) from Cambridge. This generalization was not supported in recent re-analysis of the same data using fPCA and unsupervised hierarchical clustering [10], however, which instead suggests that inner-city London English and SSBE may differ in the *realization* of a shared set of intonational categories. This prior work has largely analysed only scripted speech data, though [10] offers a first overview of contours in unscripted speech, noting an additional rise-fall contour to add to [8]'s inventory.

Interestingly most, if not all, of the anecdotal MLE features reported above (e.g., use of falsetto, raised pitch and segmental lengthening to mark emphasis) are also reported for the intonation of early varieties of African American English [11] and Eastern Caribbean Creoles [12]. Specifically, the cited creole varieties are reported: a) to differentiate statements and yes/no questions only via intonation, due to lack of auxiliary-inversion for yes/no-question formation [cf. also 13], and b) to use consonantal lengthening for discourse-pragmatic emphasis, at least in Eastern Caribbean Creoles [12].

1.3. The present study

The present study is the first analysis of MLE English intonation drawing only on (semi-)spontaneous speech collected during sociolinguistic interviews. We focus on two anecdotally salient prosodic features of MLE, identified using non-prosodic criteria: 1) polar questions of different types, and 2) instances of emphatic consonant lengthening (ECL).

For polar questions, we aim to document the range of nuclear contours deployed by MLE speakers in semi-spontaneous speech, and, further, to determine whether they vary according to presence/absence of overt syntactic marking via auxiliary inversion (i.e. yes/no questions versus declarative questions). For ECL, we aim to document for MLE, for the first time, the range of consonants targeted and the range of F0 intonational contours that co-occur with segmental lengthening.

2. Methods

2.1. Data sources

Data were sourced from two projects: (1) *Linguistic innovators: The English of adolescents in London* (*Innovators*, 2004-2007) and (2) *Generations of London English* (*GLE*, 2023-2026). Data in (1) were recorded in 2005 with informants from two London boroughs (inner London *Hackney* and outer London *Havering*) and two age groups (*old* 70+ and *teens* 16-19), for a total of 72 informants. Old participants were all of Anglo descent, but teen participants were either of Anglo or non-Anglo descent (i.e. Bangladeshi, Caribbean, African, European, etc.); data were collected via sociolinguistic interviews in friendship pairs. Data in (2) were recorded in 2023 and 2024 with informants from two broad areas of London (*East London* and *South London*) and two age groups (*older* 30+ and *teens* 16-19), for a total of 144 informants. Speakers were either of Anglo or non-Anglo descent and data was collected via sociolinguistic interviews in friendship pairs, in interviews that included a GLE-customized Diapix task [14] and a wordlist task.

2.2. Data sample selection and processing

Here we draw on the spontaneous speech collected via sociolinguistic interviews of 4 teenagers (2F/2M) monolingual English speakers of Caribbean descent recorded as part of the *Innovators* project ('old' sample) and 4 teenagers (2F/2M) monolingual English speakers of Caribbean descent recorded as part of the *GLE* project ('new' sample). We decided to look at the speech of teenagers of Caribbean heritage because the only prior work on London English intonation was with this group [8-10], and we chose recordings of teens from East London because it was here that MLE was first described [2, 3]. There were 16 such speakers in the sample (10 *Innovators* and 6 *GLE*), from which we randomly selected 8 speakers for this first 'proof of concept' pilot analysis; we favored speakers in recordings where both speakers were of Caribbean heritage while ensuring also that we obtained a final sample balanced for speaker sex.

GLE data were transcribed and diarized in WhisperX [15] and converted to Praat TextGrid format, then the transcripts were manually checked by L1 English-speaking research assistants. The Innovators data were provided fully manually transcribed. All data were then force aligned using MFA [16] at word/phone level with a GLE customized pretrained British English acoustic model and pronunciation dictionary.

The first or second author listened to the interviews to find candidate tokens for prosodic analysis on non-prosodic criteria. We identified polar questions on a 'next-turn-proof' basis [17], i.e. based on a "yes" or "no" (or equivalent felicitous response) from an interlocutor in the conversation. Polar questions were then classified as a yes/no-question (ynq), if the token displays subject-auxiliary inversion, else as a declarative question (dqu). Tokens of emphatic consonantal lengthening (ECL) were identified based on auditorily salient segmental phonetic cues. We listened to six interviews (total duration 5h40mins), ranging from 41-91 minutes and with an average duration of 57 minutes. Tokens for analysis were extracted as short wav files, with their associated TextGrid excerpt, for further analysis in Praat [18].

2.3. Analysis

The first author manually annotated position of pre-accentual, accented and post-accentual syllables in the target domain, which was utterance-final in polar questions and on the lengthened word/phrase (even if non-final) for ECL, then coded all tokens for contour shape based on auditory impression. Figure 1 shows a labelled ECL token. All labels were checked by the second author (inter-coder agreement 93%). F0 was extracted in Hz at 10ms intervals through all labelled syllables and normalised by converting to ST relative to each speaker's minimum. We used mixed effects linear regression on principle components derived from landmark registered functional Principal Component Analysis (fPCA) with time-warping [19].

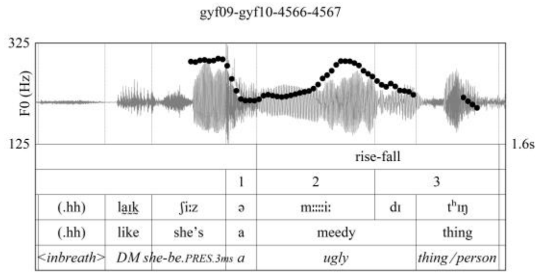

Figure 1: *F0 and waveform of sample ECL token ("meedy").*

3. Results

3.1. Overview of the dataset

We identified 206 target utterances containing a ynq, a dqu or an instance of emphatic consonant lengthening (ECL) but excluded 14 tokens due to background noise or disagreement whether the token was a clear example of the target structure. The final dataset has 192 tokens: 73 ynqs, 67 dqus and 52 ECL.

Table 1: *Token counts by type, sample and sex.*

Type	Sample	Female	Male	Total
ynqs	new	44	19	63
	old	4	6	10
	Total	48	25	73
dqus	new	31	28	59
	old	6	2	8
	Total	37	30	67
ECL	new	5	21	26
	old	5	21	26
	Total	10	42	52

Table 2 reports target domain size in terms of presence and/or absence of pre-/post-accentual syllables, by target construction. Most tokens (N=168 88%) either have all three parts of the accentual domain present (the '123' subset N=94 49%) or have a pre-accentual syllable but no post-accentual syllables (the '12' subset N=74 39%). The detailed F0 analysis in section 3.3 uses F0 trajectories from these data subsets only.

Table 2: *Domain size counts by type.*

Domain size		ynq	dqu	ECL	Total
_2_	(accsyll only)	4	4	4	12
_23	(no preaccsyll)	1	5	6	12
12_	(no postaccsyll)	25	28	21	74
123	(full domain)	43	30	21	94

3.2. Distribution of contours

Table 3 shows the count of assigned contour labels by speaker sex and sample, for ynqs, dqus and tokens of ECL. The most frequently observed contour across the whole dataset is the rise-fall. This is true for polar questions both with and without overt syntactic marking (subject-auxiliary inversion): 84% of ynqs and 64% of dqus bear a rise-fall. The next most common contour in polar questions, albeit at some distance, is the rise: 7% of ynqs, but increasing to 18% of dqus. Contours other than rise-fall are generally more commonly used in dqus than ynqs. Some of the identified polar questions may be confirmation-seeking questions rather than information-seeking questions, perhaps especially those elicited in the interactive Diapix task. The next-turn-proof criterion does not allow us to definitively differentiate information- vs. confirmation-seeking questions, but our auditory impression is that the data contains many information-seeking questions realised with a rise-fall.

The majority of ECL tokens (79%) are also realised with a rise-fall contour, followed by the fall-rise (10%). ECL tokens were much more frequently produced by males (42/52; 81%).

Table 3: *Assigned contours by type, sample and sex.*

Type	Contour	new		old		Total	%
		f	m	f	m		
ynqs	fall	2	2	0	0	4	5%
	fall-rise	1	0	1	0	2	3%
	rise	0	5	0	0	5	7%
	rise-fall	40	12	3	6	61	84%
	rise-plateau	1	0	0	0	1	1%
	Total	44	19	4	6	73	
dqus	fall	3	0	0	0	3	4%
	fall-rise	5	2	0	0	7	10%
	rise	5	6	1	0	12	18%
	rise-fall	16	20	5	2	43	64%
	rise-plateau	2	0	0	0	2	3%
	Total	31	28	6	2	67	
ECL	fall	1	0	0	2	3	6%
	fall-rise	0	0	0	5	5	10%
	rise	0	0	0	0	0	0%
	rise-fall	3	20	5	13	41	79%
	rise-plateau	1	1	0	1	3	6%
	Total	5	21	5	21	52	

3.3. The nature of the 'ubiquitous' rise-fall

Our three target constructions yielded a large majority of tokens auditorily identified as bearing a rise-fall. We thus explore the phonetic detail of the F0 contour in tokens labelled rise-fall, across construction types, to understand whether contour shape varies by construction *type* (ynq/dqu/ECL) and/or by amount of post-accentual material. For this we use only F0 trajectories from tokens which have a pre-accentual syllable, either with ('123') or without ('12') post-accentual material (see Table 2).

Figure 2 shows reconstructed curves by data subset, contour label and construction type. Due to the unscripted nature of the source data, the postaccentual domain, where present, varies in how many syllables it contains. Figure 3 zooms in to visualize contours in the '123' subset which were labelled rise-fall (N=77), by number of postaccentual syllables; the majority have either one or two postaccentual syllables: 1σ: N=49 (64%); 2σ: N=16 (21%); 3σ: N=9 (12%); 4σ: N=3 (4%).

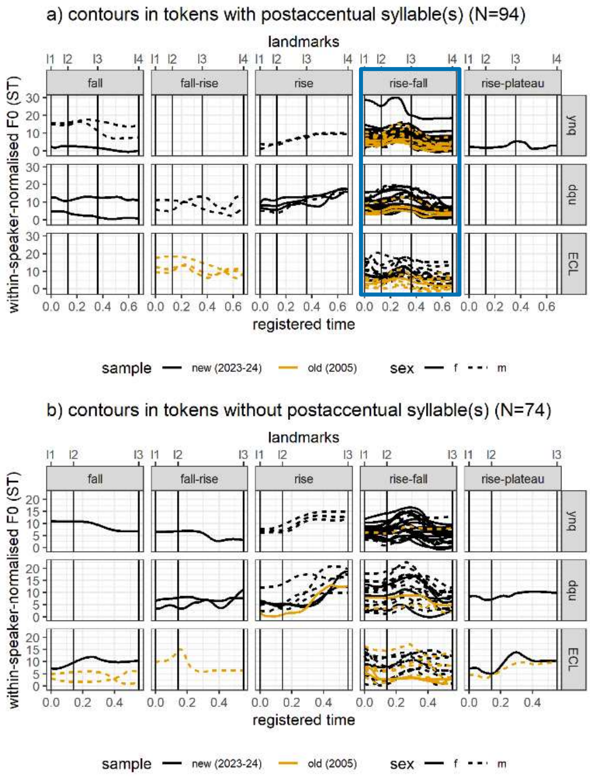

Figure 2: *Reconstructed curves in registered time for a) '123' and b) '12' subsets by type and contour label. [The blue highlighted panel is expanded in Figure 3.]*

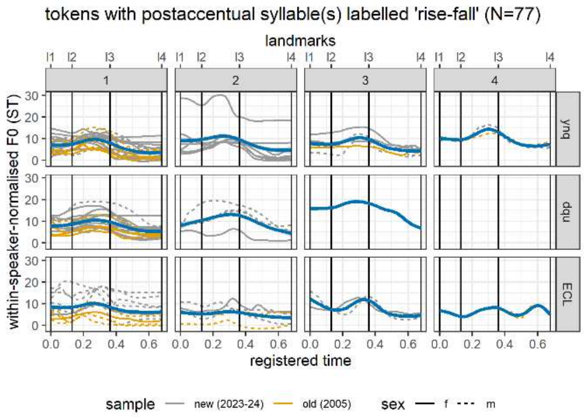

Figure 3: *Landmark registered individual curves (grey/orange) and GAM (REML)-smoothed f0 curves (blue) for tokens with postaccentual material labelled as rise-fall, by type and by nr. postaccsylls (1-4).*

We used fPCA to explore F0 contour shape in the '123' and '12' data subsets. PCs1-2 explain >90% of the variance in both data subsets, with the majority explained by PC1 (79%/75%). Adding PC3 explains only 4-5% more variance in each case, so we explore PC1-2 only. We ran mixed effects models to predict each of PC1 and PC2 in each data subset ('123'/'12'), with *type* as fixed factor and random intercepts for *speaker* and *item*, compared to a null model with random effects only. For the '123' subset, we also ran models with *number* of postaccentual syllables, speaker *sex* and/or *sample* as additional fixed factors.

We interpret PC1 in terms of overall F0 register and PC2 in terms of the alignment and scaling of F0 peaks and valleys. Figure 4 visualizes predicted PC1/PC2 by *type* for each subset. There is no effect of *type* on PC1 in the 123 subset (χ2=0.598; df=2; p=.742) but a small effect of *type* on PC1 in the 12 subset (χ2= 6.64; df=2; p=.036). In the 12 subset dqus are realised significantly higher than both ynqs and ECL, which overlap (dqu~ynq: β=-1.18; t=-2.31; p=.024; dqu~ECL: β=-1.18; t=-2.11; p=.040). For PC2, there is a main effect of *type* in the 123 subset (χ2= 10.784; df=2; p=.005), but not in the 12 subset (in a model with a random intercept for item only: χ2=2.33; df=2; p=.312); including *number* of postaccentual syllables to predict PC2 by *type* in the 123 subset does not improve model fit, nor does inclusion of *sex* or *sample*. In the 123 subset the F0 rise-fall is higher/steeper in ynqs than either dqus or ECL (dqu~ynq: β=-0.95; t=-3.34; p=.001; ECL~ynq: β=-0.73; t=-2.08; p=.040).

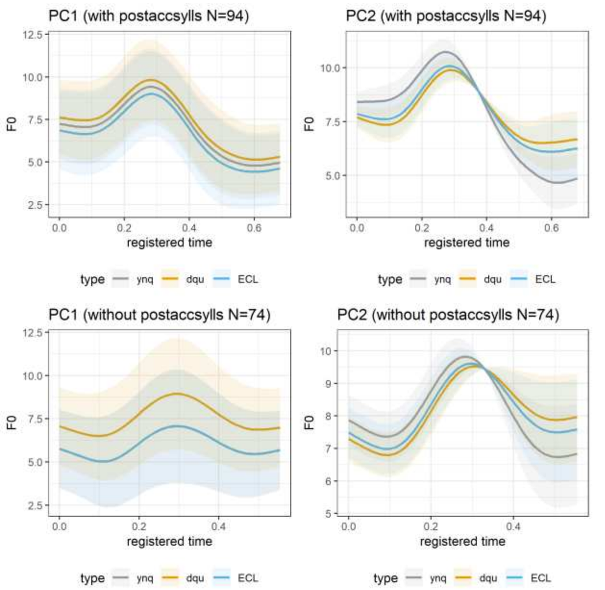

Figure 4: *Predicted PC curves by type (N=168).*

3.4. Exploring emphatic consonant lengthening (ECL)

A diverse set of lexical items carry emphatic lengthening (46 unique items in N=52 tokens), but variants of ‘bare’ (N=5), and the words ‘mad’ (N=4) and ‘mug’ (N=3), are recurrent targets. Table 4 lists which segments undergo lengthening: [m] is most frequent (21%) followed by [b], [ʃ] and [d]. Table 5 splits ECL targets by manner/place: one nasal ([m]) aside, plosives are most common; labials and coronals predominate. Targets are more often voiced (N=31; 60%) than unvoiced (N=21; 40%).

Table 4: *ECL token counts by lengthened segment.*

p	t	k	b	d	g	f	s	ʃ	m	n	l	ɹ	w	h
2	5	1	9	6	2	1	5	7	11	1	1	1	1	1

Table 5: *ECL token counts by manner and place.*

	lab	cor	dor	glottal	Total
plosive	11	10	2	0	23
fricative	1	12	0	1	14
nasal	11	1	0	0	12
approximant	1	2	0	0	3
	24	25	2	1	52

The target is word-initial in all tokens but one, which lengthens the stressed syllable onset: ‘exactly’ /ɪkˈsæklɪ/. In six tokens the target is a voiceless-obstruent-initial cluster e.g. ‘crazy’ /ˈkɹeɪzɪ/,‘prawn’ /pɹɔːn/, ‘snapped’ /snæpt/ and ‘strict’ /ʃtɹɪkt/.

Although some segment types attract ECL more frequently, it seems that almost any onset can, in principle, be lengthened. ECL was deployed much more frequently by male speakers than by female speakers, in both old and new samples. Our auditory impression is that the degree of informality and familiarity between speakers in the interview context was parallel for male and female interviewees, so this difference is unlikely to be an effect of style, though this cannot be ruled out.

4. Discussion and conclusions

The results show that a rise-fall is the most common contour in both polar questions and ECL. There were small but significant differences between utterance types in the detailed F0 contour analysis (§3.3), but we nevertheless suggest that all observed rise-fall tokens are likely to form a single phonological contour category, available to MLE speakers of both sexes across our old and new samples. From the alignment patterns in Figure 3 we would tentatively annotate the rise-fall as L+H* L%, in Autosegmental-Metrical terms [20]. The preponderance of rise-fall contours aligns with the earlier report for unscripted IViE London data [10], but contrasts with generalisations from IViE scripted data [8, 10]. We suggest that this difference is because vernacular patterns emerge more robustly in unscripted speech.

The high number of dqus in our MLE data is consistent with the observation that substrate Caribbean Creole varieties do not use auxiliary inversion to form polar questions. The use of ECL by (particularly male) speakers across both our samples is also consistent with the description of emphatic lengthening in Trinidadian Creole [12]. The segmental details of ECL in MLE merit further work, but we note here that these temporally prominent tokens (longer duration) are not cued with greater spectral prominence (expanded F0 excursion or raised register) when compared to polar questions; indeed, dqus are realised in a higher register than ECL, in one data subset. For both features, MLE speakers are using strategies available to speakers of SSBE and Mainstream English Varieties (MEV), but with a highly distinct distribution (e.g. ECL is an ‘outlier’ strategy in MEV [21]). We thus interpret both as cases of *semantic* [20] (or *distributional* [22]) variation between MLE ~ SSBE/MEV.

As for our approach, we were able to identify 192 usable targets (ynq/dqu/ECL) from six interviews for this study, so just one token per 1-2 minutes of speech. Other target features might occur more frequently and give a greater yield, and some target constructions could be partially pre-identified in transcripts (e.g. wh-questions or alternative questions, with a search for wh-words or “or”). Although painstaking, our method is a viable way of obtaining tokens of target prosodic constructions which can only be fully identified auditorily, and which are unlikely to be produced naturally (if at all) in scripted speech.

In conclusion, we have demonstrated that insights about the intonational system of a previously undescribed variety can be drawn from detailed F0 analysis of parallel tokens of target features, which have been identified in sociolinguistic interview data using independent non-prosodic criteria.

5. Acknowledgements

Funded by the Generations of London English project (UKRI ES/W013118/1) https://generationsoflondonenglish.org/.

6. References

[1] J. Cheshire, S. Fox, P. Kerswill, and E. Torgersen, "Language contact and language change in the multicultural metropolis," *Revue française de linguistique appliquée*, vol. no. 2, pp. 63–76. 2013.

[2] P. Kerswill, E.N. Torgersen, and S. Fox, "Reversing “drift”: Innovation and diffusion in the London diphthong system," *Language Variation and Change*, vol. 20, no. 3, pp. 451–491. 2008.

[3] J. Cheshire, P. Kerswill, S. Fox, and E. Torgersen, "Contact, the feature pool and the speech community: The emergence of Multicultural London English," *Journal of Sociolinguistics*, vol. 15, no. 2, pp. 151–196. 2011.

[4] S.M. Gates, *Language variation and ethnicity in a multicultural East London secondary school*, PhD Queen Mary University of London: London. 2019.

[5] P. Kerswill, A. Gibson, D. Sharma, K. McCarthy, E. Passoni, and S. Hellmuth, "Individual profiles amidst a multiethnolect: Acoustic hetereogeneity in London English," *Journal of the Acoustical Society of America*, vol. 154, no. 4_suppl., pp. A334–A334. 2023.

[6] J. Hudson, *Multicultural London English – the Urban English Accent*, Pronunciation Studio. 2020. https://www.pronunciationstudio.com/mle-multicultural-london-english-accent/.

[7] I. Thompson, *What Does Language Justice Look Like for Black British English Speakers from an Afrofuturist Lens?*, in *English: Shared Futures*: York. 2025.

[8] E. Grabe, "Intonational variation in urban dialects of English spoken in the British Isles," in *Regional Variation in Intonation*, P. Gilles and J. Peters, Editors. Tübingen: Niemeyer, 2004. pp. 9–31.

[9] F. Nolan and B. Post, "The IViE Corpus," in *The Oxford Handbook of Corpus Phonology*, J. Durand, U. Gut, and G. Kristoffersen, Editors. Oxford: OUP, 2014. pp. 475–485.

[10] E. Passoni, S. Hellmuth, and J. Pearce. "Exploring the foundations of intonational variation in (Multicultural) London English," in *Proceedings of the Third International Conference on Tone and Intonation,* 2025.

[11] D. Sutcliffe, "Gone with the wind? Evidence for 19th century African American speech," *Links & Letters*, vol. no. 5, pp. 127–145. 1998.

[12] D. Sutcliffe, "Eastern Caribbean suprasegmental systems: A comparative view, with particular reference to Barbadian, Trinidadian, and Guyanese," in *Contact Englishes of the Eastern Caribbean*, M. Aceto and J.P. Williams, Editors.: John Benjamins Publishing Company, 2008. pp. 265–296.

[13] P.L. Patrick, "Jamaican Creole: morphology and syntax," in *A handbook of varieties of English Vol. 2: Morphology and syntax.*, B. Kortmann, et al., Editors.: De Gruyter, 2004. pp. 407–438.

[14] R. Baker and V. Hazan, "DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs," *Behavior Research Methods*, vol. 43, no. 3, pp. 761–770. 2011.

[15] M. Bain, J. Huh, T. Han, and A. Zisserman, "WhisperX: Time-Accurate Speech Transcription of Long-Form Audio," *Interspeech 2023*, vol. no. pp. 2023.

[16] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger. "Montreal forced aligner: Trainable text-speech alignment using kaldi," in *Interspeech,* 2017.

[17] I. Hutchby and R. Wooffitt, *Conversation Analysis*. Vol. 2nd: Polity Press, 2008.

[18] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* 1992–2025. http://www.praat.org

[19] M. Gubian, F. Torreira, and L. Boves, "Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts," *Journal of Phonetics*, vol. 49, no. pp. 16–40. 2015.

[20] D.R. Ladd, *Intonational phonology* Second ed. Cambridge: Cambridge University Press, 2008.

[21] R. Ogden, "Making sense of outliers," *Phonetica*, vol. 69, no. 1-2, pp. 48–67. 2012.

[22] J.C. Wells, *Accents of English (Vols. 1-3)*. Cambridge: Cambridge University Press, 1982.