Exploring the foundations of intonational variation in (Multicultural) London English

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

[1]QMUL
[2]York

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

Abstract

Recent work on London English has documented a highly systematic new phonological system in inner city areas, Multicultural London English (MLE), which is argued to have arisen out of intensive, multiethnic social contact. Segmental properties of MLE have been extensively researched, but, despite anecdotal reports of an 'MLE intonation', this possibility is yet to be systematically explored. This paper revisits generalisations made in the only available description of the inventory of nuclear contours in inner city London intonation, which was based on a subset of read speech data from IViE corpus London speakers. We apply landmark registered fPCA analysis of F0 in the original IViE data subset, by the original published prosodic contour labels, and by sentence type, with a focus on yes/no questions, wh-questions and declarative questions. We also explore for the first time patterns observed in questions extracted from unscripted data from the same London speakers, based on auditory labelling of nuclear contours. The results show that the five contours listed reported for the London speakers and reflected in published IViE labels are not all differentiated in predicted curves from fPCA analysis, but instead only four distinct contours emerge. Nevertheless, the distribution of broad contour types across sentence types visualised in the fPCA results mirror those reported from the results of auditory transcription. In our exploration of contours in unscripted speech, we find a similar pattern across sentence types also, but with more use of fall and rise-fall contours.

Index Terms: intonation, nuclear contours, questions, Multicultural London English

1. Introduction

1.1. Generations of London English

Recent work on London English has documented a highly systematic new phonological system in inner city areas. This new variety, Multicultural London English (MLE [1, 2]), is argued to have arisen out of intensive, multiethnic social contact, initiating a wave of structural innovation 30 years ago [3]. The roots and trajectory of change in the complex web of London varieties are explored in the Generations of London English project [https://generationsoflondonenglish.org/].

Segmental properties of MLE have been extensively researched [1, 4], with key phonetic features ranging from unshifted, monophthongization of diphthongs [1] to TH- and DH-stopping and fronting [5], from a reduction in H-dropping and k-backing before non-high back vowels [5]. Interestingly, contemporary work [4] shows that MLE is highly stable in its segmental phonological profile, alongside subtle phonetic differences linked to community-specific ethnic distinctions. The percept of rhythm arising from these segmental properties has been described as more 'syllable-timed' in MLE than in non-MLE speakers [6]. In contrast, despite anecdotal reports of an 'MLE intonation' [7], characterized by wide pitch range and by quick and sudden pitch changes, the intonational properties of MLE are yet to be systematically explored.

1.2. The IViE London English dataset

The only available description of inner-city London intonation is the reported inventory of nuclear contours observed in read speech data from the *English Intonation in the British Isles* (IViE) project [8]. In a comparative study [9], the IViE London speakers are reported to share a core inventory of nuclear contours with speakers of Southern Standard British English (SSBE) from Cambridge (fall H*L_%, fall-rise H*L_H%, rise-plateau L*H_H%), but two further contours were used only by the London speakers (rise L*_H%, high-plateau H*_H%), and only in questions of different types. Differences in the inventory of contours could be interpreted as *systemic* variation in the sense of Ladd's taxonomy of intonational variation [10]. Grabe acknowledges, however, that whether "the different LH options for questions add nothing beyond degree of interrogativity to communicative impact [...] is an empirical question" [9 p21].

1.3. The present study

The present study revisits these generalisations using fPCA and unsupervised clustering of F0 contours in the IViE London English read speech dataset. In addition, we briefly explore for the first time patterns in unscripted data from the same speakers.

2. Methods

The IViE London data were recorded in the late 1990s with 12 teenaged monolingual English speakers of Caribbean descent from South London (6F/6M). Five tasks elicited scripted (read sentences, read story) and unscripted speech (story retelling, map task and free conversation). In the present paper we draw on the scripted read speech sentences (dec/whq/ynq/dqu) and the unscripted map task (map) and conversation (con) data. Recordings were obtained from the IViE corpus website.

2.1. Data cleaning of the corpus data

All original recordings were renamed to suit Montreal Forced Aligner (MFA) requirements for speaker adaptation and model training [10]. For example, an original recording named *j-coo1-f1.wav* was renamed to *f1_coo1.wav* (*j* indicated London in the original file-naming system but is dropped as redundant here). The read passage, story retelling and unscripted speech are made available on the IViE repository divided into chunks due to storage reasons. We therefore first concatenated consecutive chunks in Praat [11] before renaming the files from, e.g., *j-c1-m1_2.wav* to *m1_m2_con.wav* (*cl = con = free speech*). To ensure no spurious overlap after concatenating chunks, we auditorily and visually inspected spectrograms and waveforms.

Following procedures devised for the Generations of London English project [12], unscripted data was transcribed and diarized in WhisperX [13], converted to Praat TextGrid format and checked for errors by the first author. For scripted data we used published task scripts. All data were force aligned using MFA [14] at word/phone level with the GLE customized pretrained British English acoustic model and pronunciation dictionary. To compare directly with the analysis in [9] we drew on the prosodic annotation labels published on the IViE website for half of the speakers; for the London dataset we assume these are the 6 speakers analysed in [9]. The assigned contour labels were manually accessed for the subset of data in the four sentence types (dec/whq/ynq/dqu) for which 'tone' (= contour) labels are available, from www.phon.ox.ac.uk/files/apps/IViE/. Two published labels not listed in [9] (!H*L_% and !H*L_H%) were treated as H*L_% and H*L_H% respectively for analysis.

2.2. Analysis

In scripted speech, the preaccentual, accented and postaccentual syllables were manually annotated in the 4 sentence types x 12 speakers (N=202); 5 tokens were excluded due to a pause before the last lexical item. We extracted F0 in Hz at 10ms intervals through the three labelled syllables using a Praat script, plus the timepoint of four landmarks: 1) preaccentual syllable start; 2) accented syllable start; 3) accented syllable end; 4) postaccentual syllable end. F0 was 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 [15] on the time series data, to test: first, for the 6 speakers where we have IViE prosodic annotation labels (N=100), whether all five assigned contour labels are significantly differentiated; then, for all 12 speakers in the full dataset (N=197), whether unsupervised labels obtained from hierarchical clustering [16] or k-means clustering [17] support analysis in terms of five distinct contours. For each fixed factor of interest (contour label or cluster label) we ran separate mixed effects linear regression models on each of the first three principal components, comparing a model with the fixed factor and a random intercept for speaker, with a null model without the fixed factor.

In unscripted speech in the con/map tasks, we identified whq/ynq/dqu using non-prosodic criteria (i.e., presence of wh-word/syntactic inversion and/or next-turn-proof interlocutor response), yielding 36 ynqs; 41 whq; 26 dqu (N=103); their nuclear contours were auditorily labelled by first and second authors. We report counts of labelled contours by question type.

3. Results

3.1. Landmark registered fPCA of all scripted data

Time series F0 measures in the target three-syllable window (preaccentual, accentual and postaccentual), for the full scripted dataset in tokens produced without a pause before the accented syllable (N=197), were submitted to landmark registered fPCA with time-warping [15]. The first three principal components together account for 96.2% of variance; we tentatively interpret them in terms of register/scaling (PC1: 66.5% of variation), slope polarity (PC2: 23.7%) and 'wiggliness' (PC3: 6%).

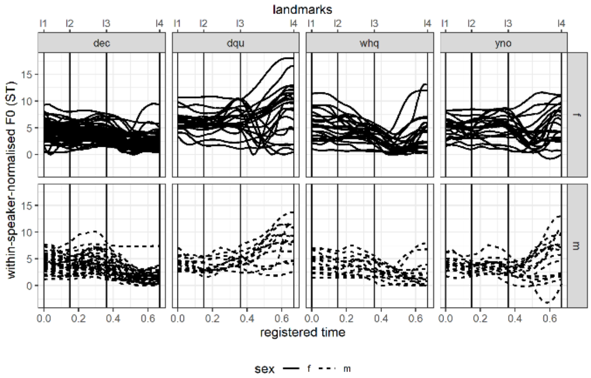

Figure 1: *Reconstructed F0 contours in registered time by sentence type and speaker sex (N=197).*

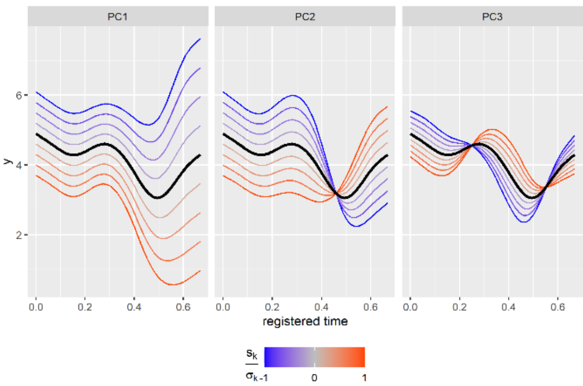

Figure 2: *First three principal components (PCs) used to model variation across observed contours (N=197).*

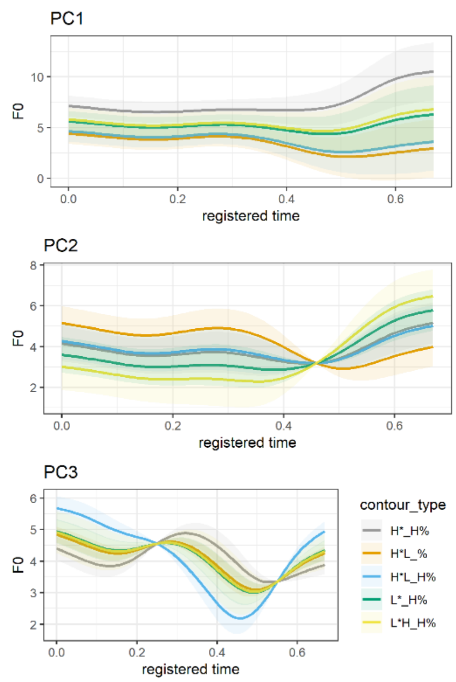

Figure 3: *Predicted fPCA curves by assigned IViE contour label in IViE-labelled data subset (N=100).*

3.2. Revisiting the IViE-labelled scripted data subset

Figure 3 shows lmer-predicted fPCA curves for each of the first three PCs by contour label in the IViE-labelled subset (N=100). PC1 does not differentiate fall H*L_% vs. fall-rise H*L_H%, nor rise L*_H% vs. rise-plateau L*H_H%. PC2 differentiates fall H*L_% vs. fall-rise H*L_H% [β=0.695; t=3.383; p=.001], but the two rising contours (rise L*_H% vs. rise-plateau L*H_H%) are undifferentiated by either PC2 [β=0.464; t=1.124; p=.264] or PC3 [β=0.023;t=0.089; p=.93]. Figure 4 shows F0 contours in registered time by assigned contour label and target sentence type for the IViE-labelled data subset. The mapping of contours to sentence type in the labelled subset matches that reported by [9] for the labelled subset, but landmark registered fPCA does not significantly differentiate as many distinct contours as envisaged in the published analysis.

3.3. Hierarchical clustering on output of fPCA

Hierarchical clustering, with Euclidean distance and complete linkage as linkage criterion following [17], was performed on fPCA scores obtained from the full scripted dataset (N=197). Figure 5 visualizes lmer-predicted fPCA curves for each of the first three PCs by hierarchical cluster label, cut at 5 clusters. PC1 significantly differentiates all clusters pairwise, except for clusters 1~2 [β=-0.175; t=-1.03; p=.3]. PC2 differentiates clusters 1~2 [β=1.039; t=8.72; p<.001] but not clusters 2~5 [β= -0.10; t=-0.174; p=.86]. For PC3, cluster 5 overlaps with all clusters, but all other clusters are differentiated from at least one other (including clusters 1~2 [β=-0.178; t=-1.996; p=.047]). Figure 6 shows a 'confusion matrix' of the reconstructed F0 contour by predicted hierarchical cluster label and by assigned IViE label, for the data subset (N=100) where we have the published IViE labels. Clusters 1, 4 and 5 each map to a single IViE label, but no IViE label maps to a single cluster label. The overlap of clusters 1~2 for PC1 (which is the PC that accounts for the majority of the variance in the data) is supported in that, although all tokens classified in cluster 1 were labelled as falls (H*_L%), many cluster 2 contours were also labelled falls.

3.4. K-means clustering on output of fPCA

We ran a base-r `kmeans` cluster analysis on fPCA scores for all data (N=197), with k=5, the default Hartigan-Wong algorithm and 25 start points. Figure 7 visualizes predicted fPCA curves for the first three PCs by k=5 cluster labels. PC1 differentiates all clusters pairwise, except 3~5 [β=0.116; t=0.415; p=.68] and 3~4 [β= 0.375; t=1.45; p=.15]. PC2 also differentiates all clusters pairwise except 3~5 [β=0.184; t=1.014; p=.312]. PC3 differentiates all clusters pairwise except 1~2 [β=0.177; t=1.912; p=.058] and 1~5 [β=-0.144; t=-1.59; p=.12]. Figure 8 shows a 'confusion matrix' of reconstructed F0 contours by predicted (k=5) cluster label and assigned label for the IViE-labelled subset (N=100). The predicted overlap of 3~5 is mirrored in the one-to-many mapping of both the 3 and 5 cluster labels to multiple IViE-assigned labels, but the same is true for all k-means cluster labels except 4 which maps only to H*L_%.

3.5. Exploration of unscripted speech

Our initial exploration of unscripted questions (Table 1) reveals a similar distribution of contours across ynqs and whqs as in read speech, though with many more falls. The realisation of declarative question (dqu) is more varied, with frequent use of a rise-fall contour not identified by [9] for the scripted data. Figure 9 below shows a rise-fall produced on an unscripted dqu.

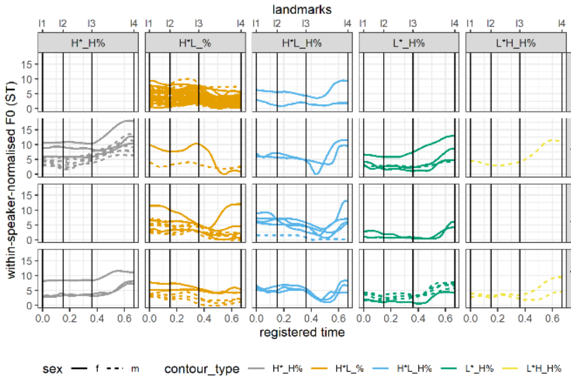

Figure 4: *F0 contours by sentence type, published label and sex in IViE-labelled data subset (N=100).*

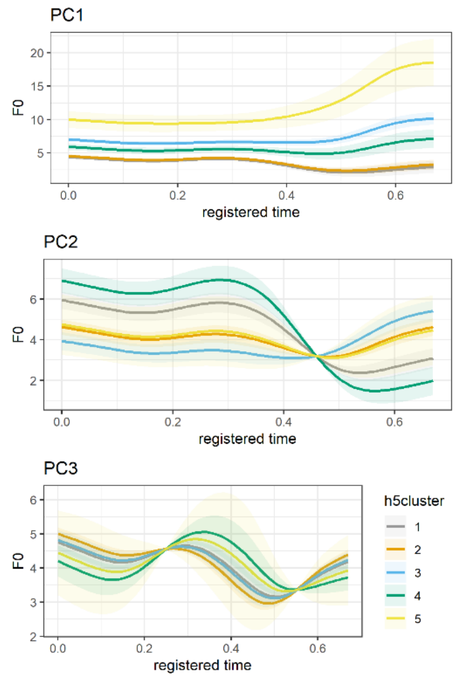

Figure 5: *Predicted fPCA curves by hierarchical cluster label (cut at 5 clusters) in all data (N=197).*

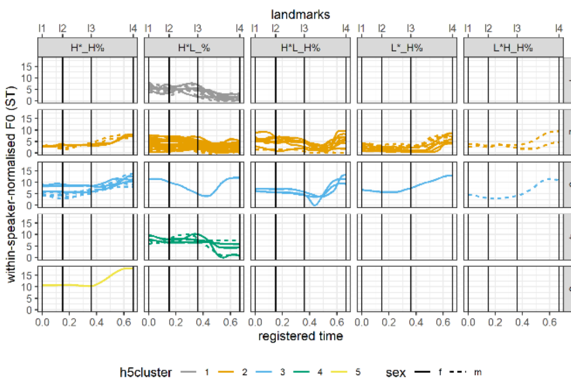

Figure 6: *Confusion matrix of contours by h=5 cluster label, published IViE contour label and sex (N=100).*

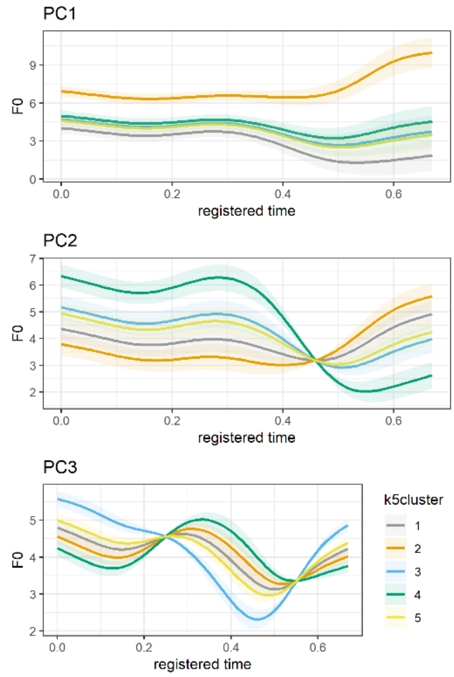

Figure 7: *Predicted fPCA curves by unsupervised k-means cluster label (k=5) in full data subset (N=197).*

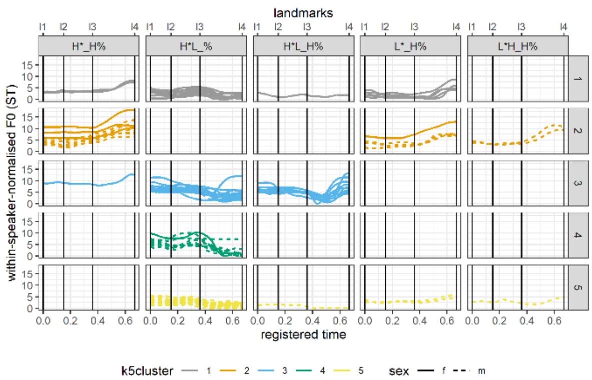

Figure 8: *Confusion matrix of contours by k=5 cluster label, published IViE contour label and sex (N=100).*

Table 1: *Contours by sentence type in unscripted speech (N=103).*

ynq	whq	dqu	contour
69%	62%	27%	fall
3%	0%	4%	fall-rise
0%	7%	4%	rise-plateau
14%	5%	12%	rise
6%	2%	0%	high-plateau
3%	14%	54%	rise-fall
9%	7%	0%	no consensus
100%	100%	100%	TOTAL

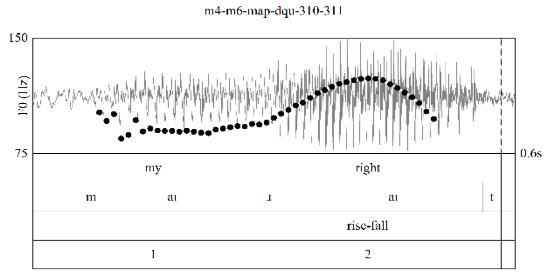

Figure 9: *Pitch trace of a sample declarative question produced with a rise-fall contour by speaker m4.*

4. Discussion

The IViE corpus London dataset provides a unique opportunity to explore the intonational phonology of 1990s London English speaking teenagers of Caribbean heritage from South London. Our starting point in this paper has been to probe the preliminary analysis of the scripted data offered by [9], not least because it hypothesized systemic differences with SSBE.

Revisiting the published IViE contour labels through the lens of fPCA shows that a model of landmark-registered F0 does not differentiate all five contours labelled as distinct in the earlier analysis, since none of the three PCs significantly differentiated the rise L*_H% vs. rise-plateau L*H_H%. Both k-means and hierarchical cluster analysis on the full scripted dataset yields a set of clusters in which all five are differentiated pairwise by at least one of the three PCs, but, for both, the match of cluster labels to published IViE labels for the labelled subset is partial at best. In contrast to [17], here the two cluster analyses differ in their mapping (to IViE labels), which may be due to a different k-means approach. Crucially, though, the rise L*_H% vs. rise-plateau L*H_H% contrast – proposed in [9] to be present in the IViE London data but not in the parallel IViE Cambridge data – is not supported by either cluster analysis.

If the spectral and temporal cues captured by landmark registered fPCA encompass all cues to the identified contours, the number of distinct contours in the data appear to be less than envisaged in [9]. However, other aspects of the signal may have informed the auditorily recognized contour distinctions, beyond the F0 and durational variation captured by the fPCA. Our own auditory annotation of patterns in unscripted data produced by the same speakers (for which no analysis has to our knowledge previously been published) points to a similar range of observed contours, even adding a further contour category into the mix.

5. Conclusion

Reanalysis of the IViE London data via fPCA and unsupervised hierarchical clustering offers an interim answer to the empirical question raised by [9], observing a distinction between only two rising contours in this variety of London English, not three. We tentatively conclude that the London English represented in the IViE London dataset (i.e., MLE) may vary from Cambridge English (i.e., SSBE) primarily in *realisation* of a shared set of intonational categories and is unlikely to display *systemic* variation (i.e. no intonational contours used in one variety but not the other). In future work we will extend the analysis to include direct comparison of fPCA modelling of IViE London and Cambridge scripted and unscripted data, as well as analysis of data from the London English Corpus which is being created in the Generations of London English project.

6. Acknowledgements

Thanks to Brechtje Post for sharing the IViE London metadata. The *Generations of London English* project is funded by the UKRI ESRC in an award to Devyani Sharma (ES/W013118/1).

7. References

[1] 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.

[2] 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.

[3] P. Kerswill and E. Torgersen, "Tracing the origins of an urban youth vernacular: Founder effects, frequency, and culture in the emergence of Multicultural London English," in *Advancing Socio-grammatical Variation and Change*. Routledge, 2020. pp. 249–276.

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

[5] J. Cheshire, S. Fox, P. Kerswill, and E. Torgersen, "Ethnicity, friendship network and social practices as the motor of dialect change: linguistic innovation in London," *Sociolinguistica*, vol. 22, no. 1, pp. 1–23. 2008.

[6] E.N. Torgersen and A. Szakay, "An investigation of speech rhythm in London English," *Lingua*, vol. 122, no. 7, pp. 822–840. 2012.

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

[8] 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.

[9] 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. Tuebingen: Niemeyer, 2004. pp. 9–31.

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

[11] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* 1992–2018. http://www.praat.org

[12] D. Sharma, P. Kerswill, K. McCarthy, and S. Hellmuth, *Generations of London English: Language and Social Change in Real Time*, UKRI ESRC. 2023–2026.

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

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

[15] 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.

[16] C. Kaland, "Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours," *Journal of the International Phonetic Association*, vol. 53, no. 1, pp. 159–188. 2023.

[17] C. Kaland, J. Steffman, and J. Cole. "K-means and hierarchical clustering of f0 contours," in *Proc. Interspeech 2024 1520-1524,* 2024.