



Data-driven approaches to pitch modelling in two Mexican Spanish ethnolects: K-means Clustering & GAMMs

Gilly Marchini¹, Jeremy Steffman¹

¹Linguistics and English Language, University of Edinburgh, UK

g.e.m.marchini@sms.ed.ac.uk, jeremy.steffman@ed.ac.uk

Abstract

This study models F0 data for documentation purposes. Using K-means clustering and Generative Additive Mixed Models, it documents differences in F0 contours between two Mexican Spanish ethnolects: Afro-Mexican Spanish and Altiplateau Mexican Spanish.

Both dialects distinguish focus condition through contour shape and pitch range: rising contours and high frequency ranges are associated with in-focus conditions, and lower frequency plateaus with post-focus. AMS however shows delayed peak alignment in open syllables and overall higher pitch contours for those with post-vocalic /s/.

We posit that these fine-grained differences in segmental alignment and pitch range would not be captured through traditional methods, e.g., Sp_ToBI. With this study's focus on phonetic detail, the ToBI model risks both the absence of phonetic variation and the assumption of underlying phonological structure. We therefore argue that this variation is best captured through clustering and GAMMs.

Index Terms: prosody, language documentation, methods, clustering

1. Introduction

This study documents the effects of focus and the segmental string on F0 contours in two Mexican Spanish ethnolects: Afro-Mexican Spanish (AFS) and Altiplateau Mexican Spanish (AMS). It considers the role of data-driven approaches (K-means clustering & GAMMs) to capturing emergent variation, and the role of this in future, phonetically-driven, documentation projects. Background on a pilot study employing Sp_ToBI to AFS is provided in §2, with the methods and parameters for analysis outlined in §3. We first analyse the clustering output for the entire dataset in §4.1, prior to variety-specific GAMMs in §4.2. We analyse the interaction between focus, post-vocalic sounds, and their syllabic affiliation in each variety. §5 discusses the appropriateness of the methods in capturing emergent variation, and the implications of this for wider documentation purposes.

2. Background

2.1. Afro-Mexican Spanish

Afro-Mexican Spanish (AFS) is an under-explored variety of Spanish spoken by isolated communities of Afro-indigenous heritage on the Costa Chica, southwestern Mexico. Documentation of the variety is scarce (see [1] for overview), with prosodic analysis limited to [2, 3]. ToBI analysis of spontaneous speech data revealed an interaction between peak alignment and the segmental string in broad focus declaratives: peaks align on

following nasals regarding of intervening syllable boundaries. Thus in /VN.C/ contexts, e.g., *descendiente* (/de.sen.ˈdjen.te/, ‘descendent’), peaks align tonically (L+H* in Sp_ToBI), and in /V.NV/, e.g., *mexicano* (/me.xi.ˈka.no/, ‘mexican’), post-tonically (L+>H*) [2, 3]. Such patterns were interpreted as nasals acting as segmental anchors, which is unexpected given the consistency of peak alignment in non-Afro Spanishes as an effect of focus, not segments [4, 5, 6, 7, 8]. Theoretically, the findings raised questions surrounding the suitability of the Segmental Anchoring Hypothesis [9, 3, 10].

Although theoretically novel, these studies are not without their methodological shortcomings. Firstly, data analysed was spontaneous speech data, meaning neither focus condition nor consonant type were controlled for. Similarly, data is from one speaker. As such, it is difficult to know whether patterns are ethnolectal or idiosyncractic, or instead common across under-explored, local Mexican varieties. Importantly, however, is the reliance on Sp_ToBI as the annotation method.

2.2. ToBI

ToBI, short for ‘Tones and Break Indices’, is a transcription schema for documenting prosody around a set of labels deemed appropriate for the language’s underlying categories and structure. The scheme is phonological in nature, framed within Autosegmental Metrical Phonology (‘AM’). Intonation is thus viewed as ‘a string of categorically distinct elements’ [9, p43]. Tones are partitioned into ‘pitch accents’ which align with metrical heads, and ‘edge tones’ which align with the metrical edges. Breaks vary in their label according to the strength of the prosodic juncture they represent, e.g., a intermediate phrase (‘ip’) or an Intonational Phrase (‘IP’) [11, 9]. The linear representation of tones and breaks encodes underlying phonological structure: tones are representative of phonological categories, and breaks prosodic structure.

ToBI initially emerged as a means to model intonation of Mainstream American English (MAE_ToBI) [11, 12], with language- and variety-specific systems having since been developed, e.g., ‘K-ToBI’ (Korean) [13], and ‘GlaToBI’ (Glaswegian English) [14]. This broad phonetic transcription is intuitive and holistic, and thus appropriate for language documentation. It is also easily acquirable and conducive to collaboration [15]. Developed to capture pan-dialectal, intonational similarities, Sp_ToBI has been successfully applied in documenting the prosody of several Spanish dialects [16, 17, 18, 4, 5, 19, 7, 8]. Pertinent to the present study is the existence of L+H* and L+>H*. Both rises with onsets in the stressed syllable, L+H* encodes tonically aligned peaks, and L+>H* post-tonic peaks. As mentioned, whilst ToBI systems have their merits and the binary rise dichotomy was useful in documenting initial pat-

terns [2, 3], we do not consider this method the most appropriate moving forwards.

Firstly, ToBI annotation is abstract. Designed for phonological analysis to capture between-category variation, not within-category, one label may encode many phonetic realisations. Here, whilst Sp.ToBI distinguishes L+H* & L+>H*, it does not specify where within the nasal syllable peaks align. It may therefore be that greater allophonic variation emerges, yet this detail is not captured. Moreover, labels do not contain information outwith the global shape, e.g., pitch range or velocity. Again, should such differences emerge between focus conditions in AFS, this would not be captured (see also [20]). Secondly, ToBI systems presuppose existing phonological knowledge of the language in question. Sp.ToBI has been constructed following Peninsular Spanish grammar [16]. Even when adapted to Mexican Spanishes [4, 21, 19], few changes have been made. Given its AM framework, Sp.ToBI works under the assumption that pitch is intonational in the variety of Spanish being analysed: it is the global pitch contour across the IP which signals phrase-level meaning. Whilst it is likely that pitch is phrasal in AFS, it is not guaranteed. Moreover, it is a contact variety, having emerged through historic and contemporary contact between African languages, surrounding indigenous languages, and Spanish. It may therefore be that complex prosodic systems emerge similar to other contact varieties (e.g., Palenquero [22] or Papiamentu [23]). Thus, applying Sp.ToBI runs the risk of not only imposing Eurocentric intonational ideals of word-prosodic systems, but also failing to capture the nuances and uniqueness of the prosody in question.

As mentioned, methodological shortcomings of [3] motivated the present study, whose interests lie in the documentation of F0 variation, and complex questions pertaining to segmental-suprasegmental interactions. Given the discussion above, it is our belief that Sp.ToBI is ill-equipped to address these. Instead, the research goals point to the need to analyse the F0 contours themselves through data-driven methods which do not presuppose underlying phonological structure. We therefore advocate for K-Means Clustering ('KM') and Generative Additive Mixed Models ('GAMMs'), which we explain in §3. We consider how methods accurately capture focus and segmental effects in peak alignment in AFS and the ambient, non-Afro Altiplano Mexican Spanish ('AMS'). We ultimately seek to contribute to an understanding of F0 analysis methods in capturing phonetic variation within and between varieties, and the implication of this for future documentation of under-researched language.

3. Methodology

3.1. Data collection

Experimental data was collected through fieldwork trips. Speakers of AFS self-identified as afro-indigenous ('*moreno*') from the Costa Chica (southwestern Mexico), and speakers of AMS European-indigenous ('*mestizo*') from Toluca (central highlands). 11 participants from each group read aloud 12-syllable sentences containing target words in in- and post-focus conditions. The target syllable contained the tonic vowel /a/ preceding /l/, /n/, or /s/ (C1). 12 word pairs were selected: 6 with post-vocalic /s/, and 6 with post-vocalic /n/ or /l/. Syllable aperture was manipulated through the inclusion of /t/ (C2), e.g., *manto* (/ˈman.to/, 'cloak') vs. *mano* (/ˈma.no/, 'hand'). Target words were randomised and received the 3rd and penultimate pitch accent. The recordings were made on a Marantz PMD661 MKII recorder with a head mounted microphone placed 12 cm

from the mouth. Recording levels were monitored throughout.

The variable context was outlined as the tonic vowel and post-tonic syllable, e.g., /a/ & /no/ for '*mano*', and /an/ & /to/ for '*manto*'. Onsets of the tonic syllable were excluded due to variability in voicelessness. F0 was sampled at 20 time-normalised intervals within each syllable with F0 trajectories z-scored within speaker. With three repetitions per speaker, 1,467 contours were analysed for AFS, and 1,523 for AMS (N = 2,990). Tokens were coded for *focus* ('post-focus', 'in-focus'), *following sound* ('sibilant', 'nasal', 'lateral'), *syllabic affiliation* ('open', 'closed'), *speaker* ('S1' - 'S22') and *variety* ('afs', 'ams'). Trajectories were then analysed through K-means clustering and GAMMs.

3.2. Clustering & GAMMs

Cluster analysis, as a general approach, is a bottom-up, unsupervised method, in which a dataset is partitioned into groups (clusters) of observations along one or more dimension. It allows researchers to make distinctions in the data without input or expectation of where these should be, thus removing the likelihood of imposing pre-conceived labels. Methods vary in how data is grouped. Specifically, K-Means clustering ('KM') partitions data around *centroids*. Clusters are identified iteratively through assigning observations based on proximity to seeds (initially) and then cluster centroids in the n-dimensional clustering space. Data points which lie closest to each centroid, calculated here through Euclidean distance, are then assigned to the cluster. This process iterates until the clusters stabilise [24, 25]. One important consideration for KM is that the number of clusters is pre-determined at the initiation of the algorithm. In an exploratory context, as with the present study, one approach is to run the algorithm with a *range of cluster numbers* that represent probable minima and maxima for groupings in the data, and then select among the various cluster solutions by some evaluation metric. This approach, as implemented in [26], is the one we take here.¹

Whereas clustering offers a descriptive, unsupervised analysis of the groups within the dataset, GAMMs require the researcher to specify the metrics along which data should be divided. As with Linear Mixed Effects Regression models, they present predictive realisations of the data, here F0 trajectories. Where confidence intervals differ may indicate locations of differences in the contours. Whilst GAMMs may be used inferentially to model F0 differences between speaker groups (e.g., [29]), we opt to include them as a descriptive tool with clustering. Clustering identifies where general distinctions emerge on global F0 contours both within and between speaker groups, and GAMMs to offer a more qualitative analysis of these.

The datasets were combined and aggregately passed through a clustering script using kml [26], and modified from the script used in [25] (available on the OSF) to visualise the mapping of emergent clusters to variety and/or focus. A range of 2:6 clusters were tested. To assess the predictive strength of emergent cluster labels to focus conditions and variety, logistic regression models were run on the dependent variable, *cluster*, (ref: *cluster A*) with an interaction between *focus* (ref: *post-focus*) and *variety* (ref: *afs*) using the lme4 package [30] with *speaker* as a random intercept. Based on these initial dis-

¹Although other clustering methods exist, e.g., Hierarchical Clustering, pilot analysis of AFS confirmed KM's ability to deal with extreme data points [24, 27, 28, 25]. Given the unexplored nature of AFS prosody, removing data could lead to the removal of important, previously undocumented variation, such that KM was selected.

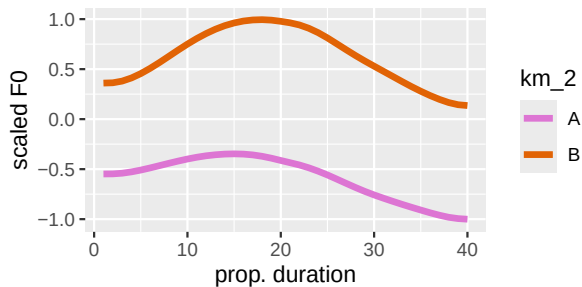


Figure 1: Mean trajectory for each cluster.

tinctions, variety-specific datasets were analysed with GAMMs, with concentration on focus distinctions and segmental, allophonic variation.

4. Results

4.1. Clustering

A 2-cluster solution was considered as best fit for the aggregate dataset as determined by the Calinski-Harabatz index [31].² Similar contours emerge: a rise reached in the tonic syllable before the 20th time-stamp (the syllable boundary) followed by a fall in the post-tonic syllable. Differences are observable for pitch range and velocity however. With scaled F0 values, Cluster A is overall much lower in the F0 space. The initial rise and subsequent fall, although visible, are reasonably shallow, covering little more than 0.25 scaled Hz. In contrast, Cluster B occupies a much higher F0 range. The rise and fall are also much steeper, covering a much greater pitch range (approximately 0.5 scaled Hz) in a similar time frame to Cluster A (Figure 1).

With clusters plotted proportionally for variety, both speaker groups show relatively equal use of each contour, with slightly higher proportional use of Cluster A. Rather, as shown in Figure 2, emergent clusters map to focus: for both AFS and AMS, Cluster A is used proportionally more in post-focus contexts, and Cluster B proportionally more in in-focus. Post-hoc pairwise comparisons reveal that across varieties, Cluster A was significantly more likely in post-focus than in-focus conditions (e.g., AMS post-focus - in-focus, $t = -24.806$, $p > .001$). Variety-specific effects also emerged: cross-variety comparisons within focus condition highlight that Cluster B was significantly more likely in in-focus and Cluster A in post-focus for AFS than AMS (e.g., in-focus AFS-AMS, $t = -2.256$, $p > .05$; post-focus AFS-AMS, 6.983 , $p > .001$). Initial cluster analysis therefore suggests that both varieties mark focus through distinctions outwith trajectory shape, with in-focus conditions marked through higher and greater pitch ranges, and steeper rises/falls. Even so, AMS makes greater use of these distinctions than AFS.

4.2. GAMMs

Syllable, sound and focus-specific GAMMs for AFS and AMS are displayed in Figure 3. Note, we do not include pairwise comparison of smooths. Instead, inferential analysis, e.g.,

²This metric, argued to be optimal as compared to other partitioning indices [26, 32], seeks to maximise between cluster variation (well-separated clusters) and minimise within-cluster variance (compact clusters), by maximising the ratio of between to within cluster variation, with normalisation for the number of clusters (see [31] for details).



Figure 2: Distribution of Clusters A and B according to focus condition for AFS (left panel) and AMS (right panel).

FPCA, is discussed in §5. For AFS, patterns show clear trajectory differences between focus conditions: those in broad focus tend to be falling or low plateaus, whilst those in in-focus are short rises followed by sharp falls, however this varies according to syllabic affiliation. Whilst few allophonic differences are observable for post-focus environments, syllabic differences emerge for in-focus contexts. In closed syllables, contours show a small, but sharp peak to a upper frequencies reached before the syllable boundary (0.50 time stamp) followed by a steep to the scaled lower range. In open syllables, nasal and lateral contours begin in the upper frequencies with a gradual fall, suggestive of a previous peak. Sibilant trajectories, however, continue to rise, reaching a maxima at the syllable boundary. Sibilant contours also tend to occupy a higher pitch range than nasals and laterals. All trajectories show sharp falls to the lower frequencies from 0.50 onwards. In all conditions, large confidence intervals occur, with much overlap with other contours, suggesting they are unlikely to reach statistical significance with follow-up FPCA.

AMS shows focus and syllable-specific patterns (Figure 3). As with AFS, differences arise for focus conditions, with post-focus trajectories as low plateaus or gradual rises at the lower frequencies of the pitch range, and in-focus sharp rises, followed by a steep fall. In both syllable types for post-focus conditions, sibilant and lateral trajectories rise gradually, whilst nasals fall or plateau. In closed syllables, peaks are reached just before (sibilant) or exactly on (lateral) the syllable boundary, whilst in open syllables, they continue rising to the end of the post-tonic vowel (time step 1.00). Differences emerge between the lateral and sibilant contours, with sibilants occupying the upper frequencies of the range, and the laterals lower. For in-focus, variation in peak alignment and pitch range is noted for syllable and segment condition. In closed syllables, peaks are reached between the 0.25 and 0.50 time steps, i.e., within C1 before the syllable boundary. In open syllables however, peaks align later, either exactly on the syllable boundary (sibilant) or just after (nasal and lateral), i.e., just on or within the onset of C1. Across both syllable conditions, sibilant trajectories occupy

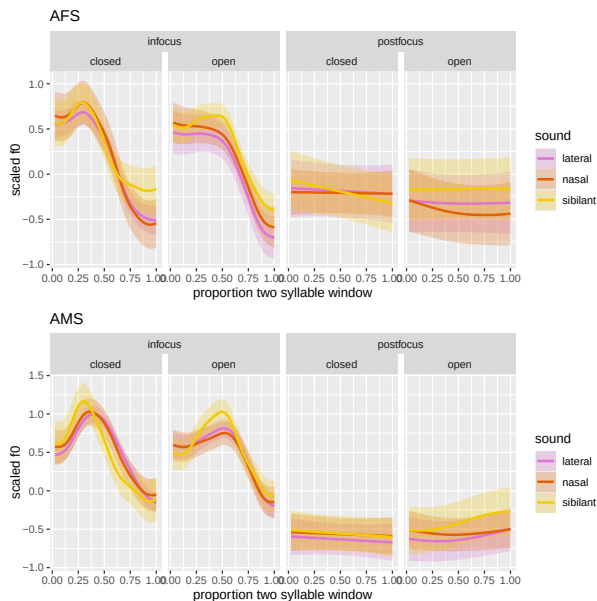


Figure 3: GAMMs according to sound (line colour) and syllabic affiliation (subset) for AFS (top) and AMS (bottom). GAMMs are subset for in-focus (left panel) and post-focus (right panel).

higher scaled frequencies, showing little overlap of confidence intervals at the onset and peak. Additionally, for closed syllables, peaks are earlier and higher, with little overlap with nasal and lateral peaks.

5. Discussion & conclusions

Combined clustering and GAMMs indicate that AFS and AMS distinguish focus through F0 contour shape and range. Post-focus conditions are signalled through low-falling plateaus, and in-focus through sharp rises and falls towards the scaled F0 maxima and minima. Dialectal difference arise, with AMS making greater use of focus distinctions and showing more segmental and syllabic variation. Namely that sibilants elicit higher frequencies, with rises in broad focus, generally higher F0 ranges, and higher and earlier aligned peaks across focus conditions. This sibilant effect is less pronounced in AFS with greater overlap with segmental positions. Moreover, differences between syllable condition in AFS are minor. In AMS however, peaks align earlier in closed syllables in both post- and in-focus contexts.

Notwithstanding the interesting theoretical implications for the Segmental Anchoring Hypothesis (see [3]), we focus on the methodological contributions for language documentation. We sought to assess the effectiveness of bottom-up clustering and GAMMs in capturing between and within variety-variation. Results indicate fine-grained phonetic differences according to speaker group, focus, segment, and syllabic affiliation. It is our belief that the use of Sp.ToBI would have both oversimplified and overcomplicated the analysis. Firstly, based on initial clustering of the aggregate dataset, both in- and post-focus trajectories would have been assigned L+H* based on their global shape, despite differences in their pitch range. A similar issue would have arisen with the assignment for labels in AMS, with sibilant contours receiving the same L+H* labels as nasal and laterals, despite GAMMs highlighting likely differences in pitch

range and peak alignment.

On the other hand, although differences in peak alignment in AFS may have been encoded through L+H* and L+>H*, this may carry an inadvertent phonological interpretation. Given the phonological nature of ToBI, assigning different labels is subliminally indicative of ascribing variation in underlying phonological categories. We are not currently in the position to conclude whether these phonetic differences constitute meaningful phonological units. Rather, we instead focus on whether (i) phonetic variation exists; and (ii) the implication of this on wider discussions surrounding variable pitch anchoring in Spanish [10]. The use of ToBI therefore simultaneously excludes key phonetic variation, and implicates a phonological conclusion that we are not in the position to make.

We thus argue that for the present study and future documentation focusing on fine-grained phonetic interactions, data-driven approaches such as clustering and GAMMs are most appropriate. These methods compliment each other: clustering is bottom-up, thus does not presuppose an underlying grammar, and is neutral to the type of F0 distinction and its temporal localisation in the contour. Nonetheless, whilst certain nuanced differences may be missed, e.g., segmental differences in peak alignment, these may be captured through GAMMs, which, although not bottom-up, is a useful tool for more detailed analyses. Clustering thus provides the metrics along which GAMMs may be run. Together, these methods also facilitate future statistical analyses. We have ascertained that notable differences in trajectories lie with regards peak alignment and frequency onset/offset. We can then use this information to inform future modelling, principally FPCA to ascertain whether indeed statistical differences emerge in these specific locations of the pitch contour.

It is important to underline that we are not denying the merit of a ToBI analysis, rather highlighting that it may be better placed later on in the documentation process. This is particularly relevant for studies such as these, which seek to document phonetic processes of language. Once these patterns have been documented through the bottom-up approaches outlined above, further experimentation, e.g., perception activities, may determine whether variation is phonetic or representative of underlying phonological categories. With this established, ToBI labels may be assigned.

In conclusion, we believe that the findings and methods above signal the utility of data-driven documentation methods. They allow one to capture complex interactions that would otherwise be missing from broad phonetic transcription without risk of imposing phonological assumptions on undocumented grammars. From the initial pitch contours, one can run further inferential analysis, such as FPCA, in order to construct a deeper understanding of fine-grained interactions. They simultaneously facilitate future, phonological and intonational modelling, which can be applied once the initial patterns are documented. We believe this method offers hope not only for the continued prosodic documentation of AFS, but also other undocumented varieties which are often overlooked.

6. Acknowledgements

We would like to thank the Costeño communities for sharing their culture and language with us. We especially thank our participants for their time and patience. Special thanks must go to Constantijn Kaland for his continued support, and to colleagues at the University of Edinburgh and the anonymous reviewers for their feedback.

7. References

- [1] N. R. Mayén, “Afro-Hispanic linguistic remnants in Mexico: The case of the Costa Chica region of Oaxaca.” Ph.D. dissertation, University of Purdue, West Lafayette, Indiana, 2007.
- [2] G. Marchini, “An exploration of pitch in Afro-Mexican Spanish,” in *ISSP 2024 Conference Proceedings*, 2024.
- [3] —, “Peak alignment in Afro-Mexican Spanish: an exploratory analysis,” in *Speech Prosody 2024 Conference Proceedings*, Leiden: Netherlands, 2024.
- [4] C. de la Mota, P. M. Butragueño, P. Prieto, and P. Fabra, “Mexican Spanish Intonation Mexican Spanish Intonation,” in *Transcription of Intonation of the Spanish Language*, P. Prieto and P. Roseano, Eds. Munich: LINCOM Europa, 2011, pp. 319–359.
- [5] P. Martín Butragueño, “El estudio de la entonación del español de México,” in *Haciendo lingüística. Homenaje a Paola Bentivoglio*, M. Sedano, A. Bolívar, M. Shiro, and A. Torres, Eds., 2006, pp. 105–125.
- [6] E. O’Rourke, “Intonation in Spanish,” in *The Handbook of Hispanic Linguistics*. John Wiley & Sons, Ltd, 2012, ch. 9, pp. 173–191. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118228098.ch9>
- [7] E. Willis, “The intonational system of Dominican Spanish: findings and analysis,” Ph.D. dissertation, 2003.
- [8] —, “Tonal levels in Puebla Mexico Spanish declaratives and absolute interrogatives,” in *Theoretical and experimental approaches to Romance linguistics*, R. Gess and E. Rubins, Eds., 2005, pp. 351–363.
- [9] D. R. Ladd, “Introduction to intonational phonology,” in *Intonational Phonology*. [Online, 2008, pp. 3–42.
- [10] P. Prieto and F. Torreira, “The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish,” *Journal of Phonetics*, vol. 35, no. 4, pp. 473–500, 2007.
- [11] M. E. Beckman and J. Hirschberg, “The ToBI Annotation Conventions,” 1994.
- [12] M. E. Beckman, J. Hirschberg, and S. Shattuck-Hufnagel, “The Original ToBi System and the Evolution of the ToBi Framework,” in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, ch. 2, pp. 9–54.
- [13] S.-A. Jun, “K-ToBI (Korean ToBI) Labelling Conventions,” 2000.
- [14] C. Mayo, M. Aylett, and D. R. Ladd, “Prosodic Transcription of Glasgow English: An Evaluation Study of Glasgow English,” *Intonation: Theory, Models and Applications*, 1997.
- [15] S.-A. Jun, “The ToBI Transcription System: Conventions, Strengths, and Challenges,” 1 2022, pp. 151–212.
- [16] M. E. Beckman, M. Díaz-Campos, J. Tevis, and M. T. A. Morgan, “Intonation across Spanish, in the Tones and Break Indices framework,” *Probus*, vol. 14, no. 1, pp. 9–36, 6 2002.
- [17] T. L. Face and P. Prieto, “Rising Accents in Castilian Spanish: A Revision of Sp-ToBI,” *Journal of Portuguese Linguistics*, vol. 6, pp. 117–146, 2007. [Online]. Available: <https://api.semanticscholar.org/CorpusID:35540926>
- [18] E. Estebas Vilaplana and P. Prieto, “La notación prosódica del español: una revisión de Sp-ToBI,” *Estudios de fonética experimental*, vol. 17, pp. 264–283, 2009.
- [19] P. Martín Butragueño, “Aproximación a la entonación del español de la ciudad de Oaxaca, México: hacia una geoprosodia,” *Moenia*, vol. 25, pp. 539–596, 2019.
- [20] D. R. Ladd, “The Trouble with ToBI,” 2 2022, pp. 247–258.
- [21] P. Martín Butragueño, “Hacia una descripción prosódica de los marcadores discursivos: datos del español de México,” in *La tonía: dimensiones fonéticas y fonológicas*, E. H. Z. and P. M. Butragueño, Eds. Mexico: Colegio de México, 2003, pp. 375–402.
- [22] J. I. Hualde and A. Schwegler, “Intonation in Palenquero,” *Journal of Pidgin and Creole Languages*, vol. 23, no. 1, pp. 1–31, 2007.
- [23] B. Remijsen and V. J. Heuven, “Stress, tone and discourse prominence in the Curaçao dialect of Papiamentu,” *Phonology*, vol. 22, pp. 205–235, 2005.
- [24] C. Kaland, “Intonation contour similarity: f0 representations and distance measures compared to human perception in two languages,” *The Journal of the Acoustical Society of America*, vol. 154, no. 1, pp. 95–107, 7 2023.
- [25] C. Kaland, J. Steffman, and J. Cole, “K-means and hierarchical clustering of f0 contours,” in *Interspeech 2024*, Kos, Greece, 2024, pp. 1520–1524.
- [26] C. Genolini and B. Falissard, “KML: k-means for longitudinal data,” *Computational Statistics*, vol. 25, no. 2, pp. 317–328, 2010.
- [27] H. I. Abdalla, “A Brief Comparison of K-means and Agglomerative Hierarchical Clustering Algorithms on Small Datasets,” in *Proceeding of 2021 International Conference on Wireless Communications, Networking and Applications*, M. Qian Zhihong }and Jabbar and L. Xiaolong, Eds. Singapore: Springer Nature Singapore, 2022, pp. 623–632.
- [28] A. Gupta, H. Sharma, and A. Akhtar, “A COMPARATIVE ANALYSIS OF K-MEANS AND HIERARCHICAL CLUSTERING,” *EPRA International Journal of Multidisciplinary Research (IJMR)-Peer Reviewed Journal*, no. 8, 2021. [Online]. Available: <https://doi.org/10.36713/epra2013>
- [29] J. Genette and J. Verhoeven, “F0 dynamics associated with prominence realisation in children with hearing impairment,” *Stem-, Spraak- en Taalpathologie*, vol. 27, 2 2022.
- [30] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
- [31] T. Caliński and J. Harabasz, “A dendrite method for cluster analysis,” *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.
- [32] Y. Shim, J. Chung, and I.-C. Choi, “A comparison study of cluster validity indices using a nonhierarchical clustering algorithm,” in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC’06)*, vol. 1, 2005, pp. 199–204.