The Sound of Code-Switching: Prosodic Profiles of Spontaneous Spanish-English Speech

Debasmita Bhattacharya[1,*,**], *Michela Marchini*[2,*], *Julia Hirschberg*[1]

[1]Department of Computer Science, Columbia University, USA
[2]Department of Computer Science & Engineering, University of Michigan, USA
debasmita.b@cs.columbia.edu, mchini@umich.edu, julia@cs.columbia.edu

Abstract

Code-switching (CSW) has been studied via many linguistic frameworks, but little is known about the *prosody* of such multilingual speech. We ask: Is CSW prosodically different from monolingual productions in spontaneous conversation, and how are variations influenced by speaker proficiency and the multilingual properties of such speech? To answer these, we examine a large-scale Spanish-English corpus over a host of language-independent prosodic features, finding that CSW prosody differs meaningfully from monolingual prosody in both English and Spanish, with greater influence on this variation coming from speaker-specific factors. Our work is the first to apply such a language-generalizable and end-to-end model-forward approach to studying spontaneous CSW prosody, and provides novel insights into multilinguality that have potential implications for the synthesis of naturalistic code-switched speech.

Index Terms: code-switching, multilingual, prosodic analysis

1. Motivation

Speakers who *code-switch* alternate between multiple language varieties during conversation [1]. Code-switching (CSW) can be performed between or within utterances and is prevalent among the growing majority of global multilingual speakers, particularly in spontaneous, informal speech. Spoken CSW has been studied through various computational frameworks, spanning morpho-syntactic analyses [2, 3], socio- and paralinguistic investigations [4, 5], and discourse-functional studies [6, 7]. In contrast, relatively little is known about the *prosodic* nature of code-switched speech, leaving great scope for exploring this uniquely spoken dimension of CSW. While recent work has begun to study prosody in multilingual domains [8], our understanding of how CSW compares prosodically to monolingual speech in any of the involved languages remains incomplete.

Prior work on this topic has taken a *perception*-oriented approach, proposing that phonetic cues such as pitch accent, speaking rate, and voice onset time (VOT) allow listeners to anticipate switches in Spanish- and Mandarin-English [9, 10, 11]. In some cases, however, it is unclear whether these cues are the same as those actually produced differently by bilingual speakers. To address this, other studies have focused on prosodic patterns in speech *production*, showing that code-switches tend to occur across, rather than within, intonational units in Spanish-English [12, 13, 14], but these have provided limited insight into the prosodic characteristics of CSW itself.

Among studies specifically investigating the nature of CSW prosody, claims range from CSW being no different from monolingual prosody in French- and Spanish-English [14, 15], to CSW as fundamentally distinct from monolingual speech in terms of prosodic-syntactic factors [13]. Intermediate claims suggest *prosodic blending* in CSW; evidence includes the presence of intonational contours from both languages in Spanish-English utterances [9]. Further segmental work has studied hyperarticulation of French- and Spanish-English vowels during CSW [16, 17] and VOT in Greek-English [18], showing that these effects can vary by social context and speaker language dominance [18, 19]. However, the role of language proficiency on CSW prosody is less conclusive among balanced bilinguals with roughly equal proficiency across languages [20, 21].

Despite the breadth of related work, it is difficult to generalize about the nature of CSW prosody; all of the above studies, except for [13], examined only *read* speech produced by a handful of speakers, with most analyses done either over single lexical items or phonemes in isolation, sometimes within specific syntactic constraints, or over a few features from a narrow prosodic class. This has resulted in insightful but contradictory and very specialized analyses focused on elicited speech.

We propose to fill the gap in our understanding of *naturally-occurring* CSW prosody by applying a uniquely macro-level approach to studying the prosodic productions of Spanish-English bilingual speakers in a large-scale corpus of CSW. We examine the extent of variation between code-switched and monolingual prosodic profiles across both represented languages and whether any differences can be explained by speaker demographic attributes or multilingual linguistic characteristics. Specifically, we ask **RQ1:** Is there utterance-level variation across a suite of language-independent pitch, energy, and duration features between code-switched and monolingual spontaneous speech? And, **RQ2:** How is this variation influenced by **(a)** speaker proficiency and **(b)** linguistic characteristics of multilingual speech? To address these **RQs**, we build on [8] and find that CSW prosody differs meaningfully from monolingual prosody in both Spanish and English, with speaker-specific factors exhibiting stronger influence than multilingual linguistic attributes. Overall, we show how and explain why spontaneous CSW is prosodically *distinct* from monolingual speech. Our main contribution is a clearly interpretable, novel insight into the prosodic nuances of spontaneous CSW, which has important applications in naturalistic multilingual speech synthesis.

2. Method

Data. We examine the Bangor Miami corpus (BM) of spontaneous Spanish-English speech [22]. BM contains 35 hours of recorded dialogue across 56 conversations, corresponding to over 46,900 transcribed utterances exchanged between 84 adult speakers from Miami, Florida in the USA. Speakers converse informally on shared topics (e.g., family life, relation-

[*]These authors contributed equally.
[**]Indicates the corresponding author.

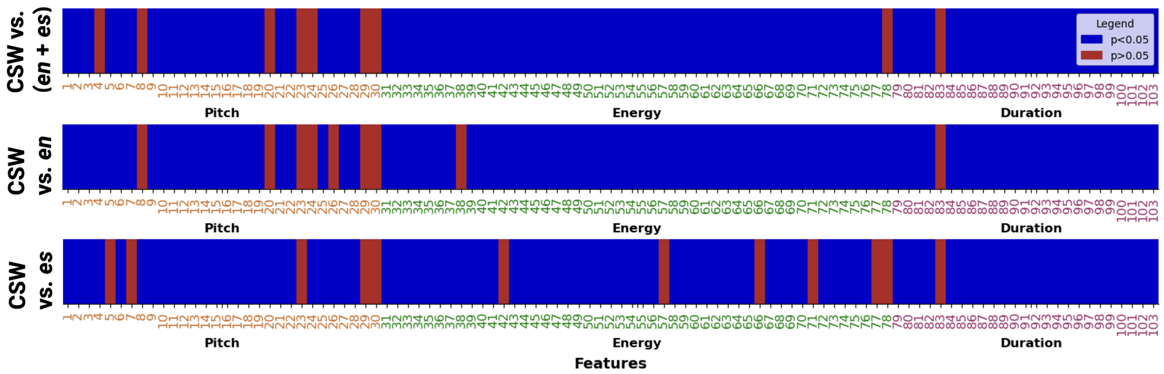


Figure 1: *p-values for prosodic feature distribution comparisons between code-switched and combined monolingual English & Spanish (top), monolingual English (middle), and monolingual Spanish (bottom) utterances. Blue indicates statistical significance; red indicates insignificance. Associated Cohen's* d *values generally fall within* $0.2 < d < 0.8$ *(i.e. small to medium effects) for significant features.*

ships, hobbies), and each produce utterances in monolingual English, monolingual Spanish, and code-switched Spanish-English, which are manually annotated with word-level language identification (LID) tags. BM also includes demographic information on each speaker, covering their language of schooling, years of language experience, self-reported ability in each language, and each of their parents' primary language. We treat these as language proficiency indicators, as done in [23].

Feature extraction. We extract and vectorize the DisVoice set of 103 utterance-level prosodic features derived from three categories – fundamental frequency (F0), energy, and duration – and six functionals[1] across voiced and unvoiced segments.[2] For consistency with prior work [8], we extract features from denoised BM data [24]. We also measure utterance-level CSW quantity and frequency using M- and I-indices [25, 26], and use prior annotations [24] of CSW direction (*en* $\rightarrow$ *es* vs. *es* $\rightarrow$ *en*) and strategy (*insertional* vs. *alternational* [27]).

Statistical testing. We compare these utterance-level prosodic feature distributions of monolingual English and Spanish to those of code-switched Spanish-English over multiple subsets of the data using feature-level independent t-tests with the Benjamini-Hochberg correction method. We support these with Cohen's d effect sizes and z-tests of proportions.

Modeling. We perform complementary 1) unsupervised and 2) supervised model analyses. For 1), we use a `scikit-learn 1.6.1` k-means clustering model with default hyperparameters. For 2), we follow [8] and use LID as a tool for understanding prosody; we treat `Whisper-base` [28] as a prosodic-LID model and evaluate its binary classification performance. This 74M parameter model was pre-trained for monolingual speech recognition in 97 languages, but allows access to intermediate LID tokens, which we restrict to English and Spanish to enforce binary evaluation. We study this model's off-the-shelf and fine-tuned LID performance using a 90/10 fine-tuning/inference split of speaker-level CSW BM data, and select the fine-tuned model checkpoint that minimizes loss.

3. Results

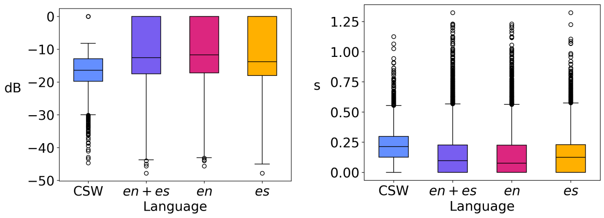
3.1. CSW differs prosodically from monolingual speech.

We first explore whether CSW prosody differs from monolingual prosody. We find that the bulk of prosodic features in code-switched utterances are significantly different ($p < 0.05$) from the same ones in monolingual English and Spanish, both when considered together and separately (Figure 1). Across comparisons, duration features drive the majority of this prosodic variation in CSW, as at least 96% of these vary significantly from their counterparts in monolingual speech. Energy features are the next most important, with no less than 87.5% of features differing significantly from corresponding monolingual speech. Pitch features contribute the least, as up to 30% of these features are similar to their monolingual analogues. These group-level patterns are the opposite of those found in [8], and help to distinguish the prosody of monolingual speech within a multilingual discourse context from that of multilingual speech itself.

We inspect these prosodic differences more closely and highlight key observations from qualitative comparisons between code-switched and monolingual speech. F0 contours of initial and final voiced segments generally have higher mean values in code-switched utterances than in monolingual ones, with greater variance in code-switched F0, particularly for initial segments. This is true relative to both monolingual English and Spanish, with consistently greater Δs between CSW and Spanish than between CSW and English. In terms of energy, initial (Figure 2a) and final voiced segments have lower mean and maximum values in CSW relative to monolingual speech in both languages. This is also true for minimum energy in unvoiced segments. Additionally, across voiced and unvoiced segments of CSW, variance in energy features is greater than in monolingual utterances. Finally, the mean and standard deviation in duration of voiced (Figure 2b) and unvoiced segments, and duration of pauses, are greater during CSW than in monolingual English and Spanish. The ratio of duration of pauses to combined voiced and unvoiced segments is also higher in CSW than monolingual speech. Cumulatively, we find that CSW tends to sound higher-pitched, quieter, and more disjointed than both monolingual English and Spanish. The consistent pattern of heightened *variance* across code-switched prosodic feature groups is striking, and echoes prior work on the diagnostic value of feature spread over absolute feature values [29].

Having found evidence of prosodic patterns that distinguish code-switched utterances from monolingual speech in both English and Spanish, we examine whether these are salient enough for a simple unsupervised model to learn and apply. Our k-means ($k = 2$; 206 params.) clustering model demonstrates reasonable ability to do so in multiple clustering scenarios, achieving an average accuracy of 85% on differentiating CSW

[1] Mean, standard deviation, maximum, minimum, skew, kurtosis.
[2] See full set at https://disvoice.readthedocs.io/en/latest/Prosody.html



(a) (b)

Figure 2: *Visualizing trends in (a) mean energy and (b) mean duration of intial voiced segments across utterance types.*

and monolingual utterances based on prosody alone (Table 1). Visual inspection of clustering outputs (e.g., Figure 3) reveals a distinct boundary between code-switched and monolingual clusters across relevant features that reinforce our findings so far. Overall, our statistical, qualitative, and unsupervised modeling results provide complementary evidence of meaningful prosodic differences between spontaneous code-switched and monolingual speech, motivating the remainder of the work.

Table 1: *k-means clustering accuracy across comparison settings. The first row is a baseline for z-tests of proportions.*

Comparison	Accuracy	p-value w.r.t. baseline
en vs. *es*	0.636	–
(*en* + *es*) vs. CSW	0.843	$p \ll 0.05$
en vs. CSW	0.837	$p \ll 0.05$
es vs. CSW	0.868	$p \ll 0.05$

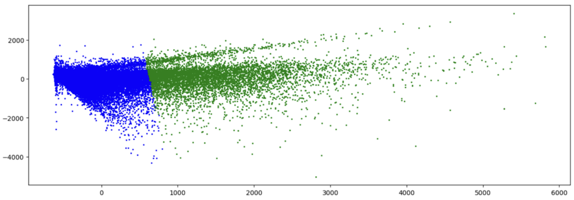


Figure 3: *Clustering all monolingual (blue) vs. code-switched (green) utterances. Axes represent the relevance of selected prosodic feature groups post-PCA; the highest loadings for PC1 (x-axis) and PC2 (y-axis) correspond to greater mean , max., and SD in energy and pitch features. Visualizations for clustering English vs. CSW and Spanish vs. CSW are almost identical.*

3.2. Prosodic variation in CSW is shaped by speaker-level language proficiency characteristics.

Having found that CSW differs prosodically from monolingual speech, we now must explain these differences. Since prior work disagrees on the influence of speaker proficiency on prosody [18, 20], we examine proficiency indicators relative to group-level prosodic features. First, we identify which language proficiency factors correspond to the greatest prosodic differences across demographic groups. Recall that demographic variables span speaker linguistic ability and language exposure factors (details in Section 2). We then apply our findings towards examining whether CSW prosody more closely resembles the monolingual prosody of speakers' higher-proficiency language, as indicated by the most salient demographic factors.

To begin, we construct two comparison groups for each language proficiency feature: for categorical features, we bin speakers according to their indicated language (*es* or *en*); for continuous ones, we assign speakers to groups based on their reported *es*:*en* ratio value (>1: *es*; <1: *en*). Then, across each pair of comparison groups corresponding to a given proficiency feature, we compare prosodic feature distributions over each of three utterance types: 1) CSW, 2) monolingual English, and 3) monolingual Spanish. In other words, in each of the three cases, we compare prosodic productions of the *same* type of utterance between speakers with opposite proficiency attributes.

We find that language proficiency factors generally correspond to prosodic differences in monolingual speech across speaker groups, but do not show consistent effects in CSW. For example, over 80% of prosodic features are significantly different across Spanish utterances produced by speakers whose secondary schooling was in English relative to those whose secondary schooling was in Spanish. The extent of differences increases to 85% over English utterances for the same comparison. Similarly, over 72% and 77% of prosodic features are significantly different across monolingual English and Spanish utterances, respectively, as produced by speakers with English-medium primary schooling relative to those with Spanish-medium primary schooling. In contrast, for code-switched utterances, the only language proficiency factor for which a majority (51%) of prosodic features differ significantly between speaker groups is years of experience. In sum, language proficiency, as defined categorically by language of schooling, appears to play a greater role in shaping prosodic variation in monolingual, rather than code-switched, speech.

Based on these findings, we hypothesize that speakers produce CSW prosody that is *more similar* to the monolingual prosody of their *higher-proficiency* language, as indicated by primary and secondary school media, than to that of their lower-proficiency language. Our experimental findings support this hypothesis. Speakers whose primary school medium was Spanish show fewer statistically significant differences between their code-switched and Spanish prosody relative to those between their code-switched and English prosody: $\Delta_{CSW/en-CSW/es} = 88\% - 72\% = 16\%$. We find a consistent pattern with respect to secondary schooling in Spanish: $\Delta_{CSW/en-CSW/es} = 89\% - 74\% = 15\%$. Similarly, considering speakers whose schooling was in English, $\Delta_{CSW/es-CSW/en} = 95\% - 86\% = 9\%$ at the primary level and $\Delta_{CSW/es-CSW/en} = 93\% - 83\% = 10\%$ at the secondary level. These Δs are all statistically significant with $p < 0.05$ per z-tests of proportions and correspond to 15-20% of baseline significant differences between monolingual English and Spanish prosodic features in each comparison, suggesting a moderate but meaningful relationship between language proficiency and the prosodic nature of CSW. This relationship is stronger among speakers with greater Spanish proficiency, tying directly into previous findings on CSW linguistic behavior and language proficiency in BM [23]. So, we have preliminary indications of speaker-level characteristics that may drive prosodic differences between CSW and monolingual speech.

We now define a LID task to assess the strength of this association: a binary classifier that can internalize the prosodic relationship between proficiency and CSW should identify a code-switched utterance by a Spanish-proficient speaker as monolingual Spanish, and one produced by an English-proficient speaker as monolingual English, given prosodic CSW inputs and only these 2 output classes. Thus, we use prosodic-LID as a proxy to identify which language CSW resembles *more* among speakers with unambiguous school-based *es* vs. *en* proficiency.

The `Whisper-base` model's off-the-shelf (baseline) per-

Table 2: *Pre-trained (PT) and fine-tuned (FT) model accuracy and class-wise F1 score on binary prosodic-LID. The fine-tuned model uses an additional classifier head (CH) in one setting.*

Model	Accuracy	F1-Score *en*	F1-Score *es*
`Whisper-base` (PT)	0.64	0.74	0.44
`Whisper-base` (FT)	0.83	0.86	0.77
`Whisper-base` (FT + CH)	0.92	0.93	0.89

formance is underwhelming on prosodic-LID for code-switched speech (Table 2), with poor generalization between language classes despite balanced test data. We fine-tune the model[3] on BM CSW data to improve LID performance, and find that it significantly outperforms its pre-trained baseline ($p < 0.01$), though performance remains modest. Such performance may be due to the model's original sequence labeling objective, rendering it imperfect for direct application to our classification task. To tackle this, we append a linear classifier head to its encoder, which leads to significant improvement ($p < 0.01$) in performance on the held-out test set of unseen speakers, both overall and for each language individually. This final set of fine-tuned results suggests that the model may well learn how CSW prosody and language proficiency are related, reinforcing the role of speaker characteristics in shaping prosodic variation in CSW and providing support for the salience of this relationship.

3.3. Prosodic variation in CSW is influenced by linguistic properties of multilingual speech.

To complete our study, we turn to the properties of CSW itself – i.e., its quantity and frequency, direction, and strategy – to explain differences between code-switched and monolingual prosody. First, we check whether the prosody of utterances containing abundant or frequent code-switches differs from monolingual speech to a greater extent than that of utterances containing sparing or infrequent code-switches. We aggregate utterance-level richness metrics (M- and I-indices) by speaker to construct groups of high vs. low-quantity and -frequency code-switchers, binning a speaker [above/below] the corpus-level median for each metric in the corresponding [high/low] group. We find that high-quantity code-switchers show greater or equal proportions of significant feature-level differences between their CSW and monolingual prosody compared to low-quantity code-switchers ($CSW/en\Delta_{high_M - low_M} = 91\% - 81\% = 10\%$; $CSW/es\Delta_{high_M - low_M} = 85\% - 85\% = 0\%$.) This pattern replicates for high- vs. low-frequency code-switchers ($CSW/en\Delta_{high_I - low_I} = 93\% - 84\% = 9\%$; $CSW/es\Delta_{high_I - low_I} = 84\% - 84\% = 0\%$), and is more notable for differences between CSW and English. Both non-zero Δs correspond to z-test p-values below 0.05, and about 11% of baseline corpus-level significant differences between monolingual English and Spanish. This pattern suggests that prosodic differences between code-switched and monolingual speech may be slightly enhanced with greater CSW richness.

Next, we examine whether prosodic differences relative to monolingual speech are more pronounced in *es*→*en* or *en*→*es* CSW. In comparisons to both monolingual Spanish and English, *es*→*en* shows slightly more variation than does *en*→*es* CSW. However, absolute differences in variation between directions (about 3%) are not significant according to z-tests, so our results fail to support a relationship between CSW direction and prosodic variation relative to monolingual speech in this corpus.

Finally, inspired by [30], we examine whether code-switched prosodic differences depend on the CSW strategy in use. Relative to English, insertional CSW differs from monolingual speech to a similar extent to alternational CSW from the same ($\Delta_{CSW_I - CSW_A} = 91\% - 87\% = 4\%$; z-test: $p > 0.05$). In contrast, relative to Spanish, insertional CSW differs from monolingual speech more than alternational CSW does ($\Delta_{CSW_I - CSW_A} = 91\% - 81\% = 10\%$; z-test: $p < 0.05$; $\equiv 17\%$ of baseline prosodic differences between the two strategies), with relevant differences concentrated in pitch and duration features. Though somewhat mixed, these results provide promising indications that CSW strategy may indeed play a role in shaping prosodic variation in CSW. Overall, among the properties studied, variation in CSW richness and strategy drive most of the differences that distinguish CSW and monolingual prosody; CSW direction appears not to play an influential role. These preliminary findings provide additional support for CSW differing prosodically from monolingual speech and contribute an additional explanatory dimension for this phenomenon.

4. Discussion

Most prior studies of prosody have focused on monolingual speech settings. The findings of those on CSW prosody have been difficult to generalize (see Section 1). Our work adds novelty, nuance, and clarity to the broad landscape of multilingual prosodic research through a language-generalizable approach to studying spontaneous CSW. From a higher level of abstraction than any individual prosodic feature, we investigate and interpret how the prosody of code-switched Spanish-English in BM differs from monolingual speech in both languages, reinforcing and scaling up prior claims of fine-grained prosodic distinctions between CSW and monolingual speech. We then demonstrate that this difference interacts with and can be explained to varying degrees by speaker proficiency and linguistic characteristics specific to multilingual speech. Further, the variation we find is salient enough to be learned by unsupervised and end-to-end predictive LID models; the latter have been particularly underutilized as diagnostic tools in prosody research, but our application of them demonstrates their largely untapped value in enabling quantitative interpretation of latent variation.

We conclude that spontaneous Spanish-English CSW is prosodically unique, with slightly greater similarity to monolingual Spanish than English in this corpus. The qualitative patterns of CSW prosody that we uncover not only offer a high-level perspective missed by prior specialized analyses of elicited CSW, but also mirror those found for group-level contributions to prosodic differences in monolingual speech [8]; we provide an additional dimension of insight across types of utterances spoken in spontaneous multilingual conversation by demonstrating how CSW prosody relates to that of monolingual speech uttered in a multilingual discourse context. Combined with [8], this enables transitive inference of how CSW compares prosodically to monolingual speech from a monolingual context, building a more complete picture of prosody in diverse speech settings. Future work could examine how our results on a single language pair generalize to other varieties including dialectal CSW, interact with interlocutor identity and entrainment, and influence real-world speech systems, e.g., by improving unnatural intonation in generated bilingual speech.

[3]All fine-tuning and inference takes < 1 hour on a T4 GPU. Fine-tuning batch size: 8; epochs: 5; learning rate: 1e-5 (FT), 5e-5 (FT+CH).

5. References

[1] S. Poplack, "Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching," *Linguistics*, vol. 18, pp. 581–618, 01 1980.

[2] C. Myers-Scotton, *Duelling Languages*. Oxford: Clarendon Press, 1993.

[3] S. Poplack, *Syntactic Structure and Social Function of Code-switching*, ser. Centro working papers, 1978.

[4] K. A. Finnis, "Variation within a Greek-Cypriot community of practice in London: Code-switching, gender, and identity," *Language in Society*, vol. 43, no. 3, pp. 287–310, 2014.

[5] D. Bhattacharya, E. Lin, R. Chen, and J. Hirschberg, "Switching Tongues, Sharing Hearts: Identifying the Relationship between Empathy and Code-switching in Speech," in *Interspeech 2024*, 2024, pp. 492–496.

[6] J.-P. Blom and J. J. Gumperz, *Social meaning in linguistic structure: code-switching in Norway*. Routledge, 1972, pp. 75–96.

[7] P. Auer, *The pragmatics of code-switching: A sequential approach*. Routledge, 2007, pp. 123–138.

[8] D. Bhattacharya, D. Sasu, M. Marchini, N. Schluter, and J. Hirschberg, "Does Context Matter? A Prosodic Comparison of English and Spanish in Monolingual and Multilingual Discourse Settings," in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Suzhou, China: Association for Computational Linguistics, Nov. 2025.

[9] P. Piccinini and M. Garellek, "Prosodic cues to monolingual versus code-switching sentences in english and spanish," in *Speech Prosody 2014*, 2014, pp. 885–889.

[10] M. Fricke, J. F. Kroll, and P. E. Dussias, "Phonetic variation in bilingual speech: A lens for studying the production–comprehension link," *Journal of Memory and Language*, vol. 89, pp. 110–137, 2016, speaking and Listening: Relationships Between Language Production and Comprehension.

[11] A. Shen, S. Gahl, and K. Johnson, "Didn't hear that coming: Effects of withholding phonetic cues to code-switching," *Bilingualism: Language and Cognition*, vol. 23, no. 5, p. 1020–1031, 2020.

[12] P. S. Shenk, "The interactional and syntactic importance of prosody in Spanish-English bilingual discourse," *International Journal of Bilingualism*, vol. 10, no. 2, pp. 179–205, 2006.

[13] J. W. Steuck, "The Prosodic-Syntactic Structure of Intra-sentential Multiword Code-Switching in the New Mexico Spanish-English Bilingual Community," Ph.D. dissertation, The Pennsylvania State University, 2018.

[14] R. Torres Cacoullos, "Code-switching strategies: Prosody and syntax," *Frontiers in Psychology*, vol. Volume 11 - 2020, 2020.

[15] F. Grosjean and J. L. Miller, "Going in and out of languages: An example of bilingual flexibility," *Psychological Science*, vol. 5, no. 4, pp. 201–206, 1994.

[16] K. Muldner, L. Hoiting, L. Sanger, L. Blumenfeld, and I. Toivonen, "The phonetics of code-switched vowels," *International Journal of Bilingualism*, vol. 23, no. 1, pp. 37–52, 2019.

[17] D. J. Olson, "The phonetics of insertional code-switching: Suprasegmental analysis and a case for hyper-articulation," *Linguistic Approaches to Bilingualism*, vol. 2, no. 4, pp. 439–457, 2012.

[18] M. Antoniou, C. T. Best, M. D. Tyler, and C. Kroos, "Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching," *Journal of Phonetics*, vol. 39, no. 4, pp. 558–570, 2011, cross-Language Speech Perception and Variations in Linguistics Experience.

[19] D. J. Olson, "The impact of code-switching, language context, and language dominance on suprasegmental phonetics: Evidence for the role of predictability," *International Journal of Bilingualism*, vol. 20, no. 4, pp. 453–472, 2016.

[20] R. K.-Y. Tsui, X. Tong, and C. S. K. Chan, "Impact of language dominance on phonetic transfer in Cantonese–English bilingual language switching," *Applied Psycholinguistics*, vol. 40, no. 1, p. 29–58, 2019.

[21] B. E. Bullock, A. J. Toribio, V. González, and A. Dalola, "Language dominance and performance outcomes in bilingual pronunciation," in *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference*. Cascadilla Proceedings Project Somerville, MA, 2006, pp. 9–16.

[22] M. Deuchar, "Miami corpus: Preliminary documentation - bangortalk," 2011. [Online]. Available: http://bangortalk.org.uk/docs/Miami_doc.pdf

[23] D. Bhattacharya, A. Tolat, and J. Hirschberg, "From Context to Code-switching: Examining the Interplay of Language Proficiency and Multilingualism in Speech," in *Interspeech 2025*, 2025, pp. 4528–4532.

[24] D. Bhattacharya, S. Ding, A. Nguyen, and J. Hirschberg, "Measuring entrainment in spontaneous code-switched speech," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 2865–2876.

[25] R. Barnett, E. Codó, E. Eppler, M. Forcadell, P. Gardner-Chloros, R. van Hout, M. Moyer, M. C. Torras, M. T. Turell, M. Sebba, M. Starren, and S. Wensing, "The LIDES Coding Manual: A document for preparing and analyzing language interaction data Version 1.1—July, 1999." *International Journal of Bilingualism*, vol. 4, no. 2, pp. 131–270, 2000.

[26] G. A. Guzman, J. Serigos, B. E. Bullock, and A. J. Toribio, "Simple Tools for Exploring Variation in Code-switching for Linguists," in *Proceedings of the Second Workshop on Computational Approaches to Code Switching*. Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 12–20.

[27] P. Muysken, *Bilingual speech: a typology of code-mixing*. Cambridge University Press, 2000.

[28] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," 2022, Model accessed from https://huggingface.co/openai/whisper-base. [Online]. Available: https://arxiv.org/abs/2212.04356

[29] D. Bhattacharya, J. Chi, J. Hirschberg, and P. Bell, "Capturing formality in speech across domains and languages," in *Interspeech 2023*, 2023, pp. 1030–1034.

[30] R. Pattichis, D. LaCasse, S. Trawick, and R. Torres Cacoullos, "Code-switching metrics using intonation units," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 16 840–16 849.