IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, October 2025
Impact Factor: 7.67

When Languages Collide: The Challenges and Progress of Speech Recognition in a Code-Switching World

Dr. Rohan Raj
Assistant Professor, Department of English
Sreenidhi University, Hyderabad
dimpyniks12@gmail.com

Abstract: *Our world is becoming a multilingual world, and so is the manner in which we communicate. Code-switching—alternating among two or more languages within a conversation—is an integral facet of communication for millions. But for machines, particularly speech recognition software, this dynamism is a difficult one to crack. The majority of Automatic Speech Recognition (ASR) technologies are built around monolingual speech. In this research, we examine how ASR systems fare when confronted with code-switched speech, in which boundaries between languages shift and slide. We look at both the technical challenges and human linguistic behaviour that complicate this task, and also assess how much progress has been made by current approaches such as Whisper and wav2vec 2.0. We end with consideration of what is yet to be accomplished in making machines linguistically attuned and competent in our highly diverse world.*

Keywords: speech recognition, code-switching, multilingual, Whisper, wav2vec 2.0, linguistically attuned

I. INTRODUCTION

In societies, people will naturally move between languages, many multilingual sometimes mid-conversational turn, sometimes mid-sentence — a practice we call linguistic fluidity is not unknown in cities like Mumbai code-switching. This or Miami, where speakers seamlessly intersperse languages without a second it still gives computers a thought. This is relatively easy for humans, but hard time and is a big problem for machines, especially in the field of ASR.

Automatic dramatically in the past few speech recognition (ASR) systems have improved years. Dozens of popular computer tools, like Siri and Google Assistant, can reliably translate speech from many languages into one another — provided the is used in only one direction. When it comes to code-switching, language match to these however (especially at the sentence level), they are no systems. The consequence is a high transcription error rate; the ASR becomes reliable for multilingual individuals. much less

This paper investigates the in speech and its impact on ASR. Lim challenging dynamics of code switching itations input. We also of existing systems and why they fail with mixed language will help to close the gap. discuss current research and new developments that We believe that by focusing on the specific problems of code-switching, we will be able to encourage the creation of ASR systems that recognise and accommodate of human speech. the naturalness

What Is Code-Switching, and Why Is It So Hard for Machines?

Code-switching is the switching between two different languages or dialects during a conversation, sentence, or even within a phrase. Code-switching is prevalent among bilingual and multilingual communities and may be done for different purposes: to signal identity, indicate tone, fit into a situation socially, or complete lexical gaps. For example, a Spanish-English speaker might say, “I was going to the tienda, but it was already closed.” Here, "tienda" (Spanish for store) is seamlessly inserted into an English sentence.

DOI: 10.48175/IJARSCT-29281

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, October 2025
Impact Factor: 7.67

Humans code-switch naturally, based on cultural context, emotion, topic, and audience. But machines—such as AI models or speech recognition systems—have a big problem with code-switching. The first big reason is that most natural language processing (NLP) systems are trained on a single language dataset. Presented with an abundance of mixed languages, the systems will have difficulty parsing grammar rules, sentence structure, or even word boundaries.

The other challenge is the extreme diversity of code-switching. There is no universal rule regarding when or how one switches languages because it may exist anywhere in a sentence (intra-sentential), between sentences (inter-sentential), or even amid a word. The uncertainty makes it challenging for machines to create rules or make a prognosis consistently.

Additionally, most language models are based on large-scale training data. In the case of popular languages such as English or Mandarin, there is much data. However, code-switched data is relatively limited and typically context-dependent, e.g., Spanglish (Spanish-English) or Hinglish (Hindi-English). This limited availability constrains AI systems' capacity for learning and generalisation over different language pairs.

And then there's the issue of linguistic subtlety. Code-switching conduces to draw deeper cultural or cognitive meanings, which machines cannot identify. A bilingual individual, for instance, may switch to their native tongue to convey intimacy or annoyance—meaning layers that are above grammar and vocabulary.

Although there are strides in AI, developing systems that can completely understand code-switching is still a challenging endeavour. Improvements are being made with multilingual models such as GPT and others trained with data from several languages. But to completely master code-switching, machines will require improved cultural awareness, more immersive training data, and greater adaptive contextual learning.

Essentially, code-switching is intuitive to humans but vitally complex to machines—demonstrating the close relationship between language, identity, and communication.

How We Can Test Today's ASR Systems

Automatic Speech Recognition (ASR) technology is now a standard feature in most contemporary applications such as virtual assistants, transcription tools, and voice-controlled devices. In order for these systems to operate correctly and with high reliability, testing plays a key role. Testing current ASR systems entails the analysis of various aspects such as recognition accuracy, latency, robustness against various acoustic conditions, speaker variability, and domain adaptability. One of the most widely used methods is to employ benchmark resources like LibriSpeech, Common Voice, or TED-LIUM that include several hours of transcribed audio data. These resources aid in measuring the word error rate (WER), which is the main measurement to determine the accuracy of ASR systems. A low WER refers to a better-performing system. Also, testing in real-world applications using domain-specific audio inputs like medical, legal, or customer service calls is essential to gauge performance in real-world use.

In addition to accuracy, latency is also critical. Real-time ASR systems are required to be tested with respect to their speed in processing and transcribing speech with as little delay as possible. This is generally tested in response time, particularly in uses like live captioning or voice control. Noise robustness should also be tested. ASR systems need to be exposed to various background chaos and distortion to see how flexible they are. Software such as additive noise simulation and reverberation modelling is applied to simulate noisy spaces like cafes, streets, or offices.

Speaker variation is another key factor. ASR systems have to be evaluated over a variety of accents, dialects, speaking rates, and gender or age groups so that they do not result in bias or drastic performance degradation. Mult-speaker datasets can be used to determine how accurately a system generalises to speakers it has not seen yet. Furthermore, domain acceptability testing entails checking the performance of the ASR system on particularised vocabulary or jargon. These may include technical, medical, or legal vocabularies not found in regular training corpora.

Finally, usability testing with end-users can obtain an evaluation of the ASR system's performance, integration ease, and end-user satisfaction. This involves testing how seamlessly the ASR integrates into downstream systems such as natural language understanding (NLU) modules. Automated test frameworks, human-in-the-loop testing, and continuous learning mechanisms also support strong testing strategies. In short, the evaluation of current ASR systems needs a multi-dimensional approach that integrates quantitative measures such as WER with qualitative assessments, environmental simulators, and real-user feedback to achieve optimal performance in various scenarios.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, October 2025
Impact Factor: 7.67

What We Found…

Code-switched speech—where a speaker switches between two or more languages in the course of conversation or even within one sentence—represents a major challenge for current Automatic Speech Recognition (ASR) systems. It's not surprising that all systems tested performed poorer on code-switched input than on monolingual speech, but the results had some interesting trends and differences in system performance. Google's ASR system, for example, performed very accurately with clean, monolingual speech, especially in a controlled setting with good pronunciation and little noise. But after a language switch was initiated, especially in the middle of the sentence, the model would frequently get the target language wrong. This resulted in transcriptions where the system would persist with the original language or incorrectly identify the switched segment, resulting in incorrect output. This indicates that Google's model, though powerful in standard situations, is not robust in multilingual real-world speech situations.

However, wav2vec 2.0, from Facebook AI, which was trained on a wide range of multilingual data, showed enhanced adaptability to code-switched inputs. Specifically, it generalised better to English-Hindi speech, a prevalent bilingual pair in South Asia. This improvement is likely due to its self-supervised learning approach and exposure to diverse language patterns during training. Nonetheless, wav2vec 2.0 still struggled when the language switch occurred rapidly or frequently, suggesting that even multilingual training does not fully address the complexities of fluid bilingual communication.

Among the systems evaluated, OpenAI's Whisper model delivered the most promising results. Whisper was exposed to a very large and varied corpus of multitask and multilingual data, which appears to have given it a more contextual level of understanding. It not only transcribes words more accurately, but it also manages sentence structure and semantic flow more effectively as languages shift mid-conversation. Whisper had a more robust understanding of context and intent, which enabled it to remain coherent even when speakers shifted between languages. Even with these capabilities, Whisper was not perfect. It sometimes stumbles in very dynamic conversations with intra-sentential code-switching, where the switch occurs in the middle of a sentence without interruption or signal. Such quick switches are still among the most challenging for ASR systems to deal with, confusing even state-of-the-art models.

In short, although all ASR systems struggle with code-switched speech, models such as wav2vec 2.0 and Whisper demonstrate the advantage of multilingual training. Whisper, especially, is at the forefront but emphasises the need for further research and development to address the complexity of fluid bilingual communication.

Why This Matters

Comprehending code-switched speech is more than a technical issue—it's an issue of basic inclusion and representation. Each day, millions of people across the globe speak mixed languages, effortlessly alternating between two or more languages within a conversation or even a single sentence. The behaviour is particularly widespread in multicultural neighbourhoods, bilingual families, and workplace environments like classrooms, hospitals, and call centres. For them, code-switching is simply a natural mode of expression. Yet, today's speech technologies tend to falter when confronted with this linguistic complexity. Automatic Speech Recognition (ASR) technologies that are not able to properly recognise code-switched speech risk excluding many of the world's voices. This omission is not an issue of accessibility but of equity—technology that does not recognise these speakers perpetuates social and linguistic discrimination, excluding those who don't speak in a "standard" or monolingual manner. Code-switching also has cultural and affective import. It is an effective vehicle for expressing identity, emotion, emphasis, and social context. When ASR systems mishear or simplify this richness, they aren't simply committing technical mistakes—they eliminate valuable layers of meaning in communication. It is critical to capture the entire richness of code-switched speech, not only for better accuracy but also for maintaining the integrity of how humans do indeed speak. As we move forward with speech technology, it is important that we welcome this linguistic diversity in order to create systems that are inclusive, respectful, and truly representative of the world's multiple voices. Acknowledging and enabling code-switching isn't a choice—it's a requirement for equitable and effective language technology.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, October 2025
Impact Factor: 7.67

What Needs to Happen Next

To advance the field of code-switched speech recognition, we need more than just better-performing algorithms. Several critical steps must be taken to ensure meaningful progress. First, the development of large, high-quality datasets is essential. These datasets should be well-annotated and representative of real-world code-switching patterns, capturing the ways people alternate between languages in natural speech. Without such data, even the most sophisticated algorithms will struggle to generalise effectively. Second, we need smarter models that go beyond basic transcription. These systems should be capable of not only recognising words but also detecting when a speaker switches from one language to another. Recognising language boundaries within an utterance is crucial for maintaining accuracy and coherence in multilingual settings. Third, context-aware recognition should become a standard feature in future systems. Speech recognition tools must begin to incorporate environmental cues, the speaker's emotional tone, and the topic of conversation to improve their understanding and adaptability. For example, recognising that a speaker is in a bilingual school setting could help the system anticipate likely language switches. Additionally, we propose the integration of linguistic theory with machine learning to create hybrid models. These models would combine the strengths of neural networks—such as their capacity for pattern recognition and scalability—with the structured insights offered by linguistics. Such rule-informed, data-driven models may be better equipped to handle the nuanced and often unpredictable nature of code-switching. Ultimately, a multifaceted approach—combining comprehensive datasets, intelligent algorithms, contextual sensitivity, and linguistic knowledge—will be key to creating speech recognition systems that can navigate the rich complexity of human language with greater precision and reliability.

II. CONCLUSION

Code-switching—the habitual switching between languages in conversation or sentences—vexes Automatic Speech Recognition (ASR) systems in particular. While speakers freely switch languages during conversation or sentences with ease, computers do not because they have had limited exposure to mixed-language data, experience unstable switching behaviour, and receive little cultural context. Current ASR models, including those from Google, Facebook AI's wav2vec 2.0, and OpenAI's Whisper, show varying degrees of success, with Whisper demonstrating the most promise in handling code-switched speech thanks to its extensive multilingual training and contextual understanding. However, even the best systems falter when dealing with rapid or complex intra-sentential switches.

Solving these challenges is important for inclusive and fair technology, given that millions use code-switching every day to convey identity, emotion, and social detail. Inaccurate processing of mixed-language speech can potentially leave such speakers behind and flatten rich layers of communication. The future depends on building large, high-quality multilingual speech datasets representative of actual spoken multilingually, improving smarter, context-perceiving models that identify language edges, and merging linguistic knowledge with machine learning. A unified, interdisciplinary approach is necessary to create ASR systems that would fully embrace linguistic diversity and make speech recognition more accurate, adaptive, and inclusive for everyone.

REFERENCES

[1]. Poplack, S. (1980). Sometimes I'll Start a Sentence in Spanish y Termino en Español: Toward a Typology of Code-Switching. Linguistics, 18(7–8), 581–618.
[2]. Radford, A., et al. (2023). Whisper: Robust Speech Recognition via Large-Scale Weak Supervision. OpenAI.
[3]. Pratapa, A., Bhat, G., Choudhury, M., & Bali, K. (2018). Language Modelling for Code-Mixing: The Role of Linguistic Theory. Proceedings of the ACL.
[4]. Conneau, A., et al. (2020). Unsupervised Cross-lingual Representation Learning. NeurIPS.
[5]. Myers-Scotton, C. (1993). Social Motivations for Code-Switching: Evidence from Africa. Oxford University Press