frontiers | Frontiers in Computer Science

TYPE Original Research
PUBLISHED 12 August 2025
DOI 10.3389/fcomp.2025.1565604

OPEN ACCESS

EDITED BY
Yinghao Li,
ByteDance Ltd., United States

REVIEWED BY
Jing Shao,
Hong Kong Baptist University, Hong Kong SAR, China
Silvan Mertes,
University of Augsburg, Germany

*CORRESPONDENCE
Tyler Méndez Kline
✉ tmendezkline@berkeley.edu

RECEIVED 23 January 2025
ACCEPTED 28 July 2025
PUBLISHED 12 August 2025

CITATION
Méndez Kline T and Zellou G (2025) The perception of code-switched vs. monolingual sentences in TTS voices.
Front. Comput. Sci. 7:1565604.
doi: 10.3389/fcomp.2025.1565604

The perception of code-switched vs. monolingual sentences in TTS voices

Tyler Méndez Kline[1]* and Georgia Zellou[2]

[1]Department of Linguistics, University of California, Berkeley, Berkeley, CA, United States, [2]Department of Linguistics, College of Letters and Science, University of California, Davis, Davis, CA, United States

This study examines the intelligibility of English and Spanish lexical items in code-switched utterances across different text-to-speech (TTS) synthesis methods. Using stimuli generated with neural and concatenative TTS, 49 Spanish-English bilingual participants listened to 96 sentences, mixed with noise, and typed the phrase-final keyword. Half of the sentences contained English-Spanish code-switches (equal number of English and Spanish target keywords), and half were monolingual sentences (half English, half Spanish). Accuracy was coded binomially for correct word identification. Results show that intelligibility is lower: (1) when the target words are produced in Spanish, and (2) in code-switched conditions. These results are in contrast with previous work showing intelligibility differences between TTS conditions. Moreover, the lower intelligibility results in the Spanish target word sentences and code-switched conditions present motivations for improving voice-AI speech to include common bilingual practices.

KEYWORDS

code-switching, voice-AI, human-device interaction, switch costs, bilingualism

1 Introduction

We are in a new digital era; the use of devices like Siri and Amazon Alexa are becoming more common in daily life. These devices are used to complete various tasks (e.g., creating shopping lists, setting reminders, etc.) as well as to help with information retrieval. Their growing use is being considered among researchers interested in seeing how voice-AI fares in spaces like classrooms and healthcare settings (Kim et al., 2022; Zhan et al., 2024). The fact that many of these devices require human speech recognition and produce highly human-like speech patterns motivates inquiries about how language users communicate with such systems. Some work has begun to investigate this realm, demonstrating ongoing effects in both production and perception. Additionally, this motivates more work investigating how diverse types of linguistic modes interact with voice-AI, including bilingualism. The current study examines effects of voice-AI code switching on language processing by Spanish-English bilingual users.

Bilingualism is a ubiquitous reality for many communities across the globe, in which speakers communicate daily in multiple varieties, such as Spanish and English. In the US alone, data from the Census Bureau estimates that over 43 million, or roughly 13% of the American population, report some level of proficiency in Spanish (U.S. Census Bureau, U.S. Department of Commerce, 2023). Although English continues to be the main language used across institutional settings, the presence and growing rate of multilingualism in the US continues to motivate researchers to consider linguistic diversity in voice-AI use and development. A key aspect of this is supporting *code-switching*, a common linguistic behavior in many bilingual communities.

While the term *code-switching* has been described using a range of different formulations depending on disciplinary perspective (Nilep, 2006; Mabule, 2015), here we define code-switching as the use of more than one variety or language in the same conversational context or utterance, based on previous literature (Auer, 1984; Heller, 1988). This could include different types of switches (Poplack, 2000; Mabule, 2015), such as:

1. Extrasentential switches, consisting of tag word insertion from another variety. Example: You went to the store yesterday, verdad? *You went to the store yesterday, right?*
2. Intersentential switches, in which speakers switch between varieties at sentence boundaries. Example: She works in programming y lo hace bien. *She works in programming and she does it well.*
3. Intrasentential switches, in which speakers switch between varieties within the same sentence/utterance. This is the switch type we utilize in our study. Example: Él tiene una historia for everything. *He has a story for everything.*

In many Spanish-speaking communities across the U.S., code-switching is a very common practice that represents the linguistic norms of bilingual speakers (Barker, 1947; Cooper, 2013). Considering its ubiquitous nature, integrating natural linguistic patterns like code-switching into text-to-speech (TTS) models can have many practical applications and user benefits, such as greater accessibility to content that bilingual speakers consume more regularly, like mixed-language media, online resources, or using TTS to read digital sources. Additionally, incorporating code-switching into TTS models could help improve user experience by having more natural and realistic speech technology for multilingual users. This can enhance the usability of devices that rely on TTS systems, promoting increased human-device interaction due to having devices that match the user. Finally, incorporating code-switching into TTS voices could enhance cultural sensitivity in technology design. Voice-AI devices that pronounce cultural terms and phrases using the same phonetic and prosodic patterns that bilingual speakers use across their varieties can further enhance user experience by promoting and representing natural linguistic diversity.

While voice-AI enabled devices were originally designed to be monolingual, many different devices have been recently updated to permit multiple varieties in the same context to some extent, including Google Assistant and Amazon Alexa. However, there is a dearth of resources to incorporate different varieties and a lack of code-switching capabilities in both speech generation and recognition across different voice technologies. This disparity warrants further work to assess the current state of digital devices and explore how this technology can be developed to be more linguistically inclusive (Cihan et al., 2022; Kann, 2022). Some work has begun to develop automatic speech recognition (ASR) tools that permit the use of multiple varieties at once, such as Google Assistant, though code-switching remains to be fully incorporated (Toshniwal et al., 2018). Code-switching is an important phenomenon to examine in both the realm of production and perception.

One question that remains unanswered is how code-switching affects user performance in intelligibility. The current study focuses on this question. Previous studies have explored the perceptual effects of switching on listeners' intelligibility performance in human-human interactions. For example, Piccinini and Garellek (2014) found that Spanish-English bilingual listeners rely on prosodic contours to anticipate upcoming switches, demonstrating how integral linguistic cues are for comprehension. Similar studies investigate related effects, such as the absence of phonetic-level cues creating switch costs (Shen et al., 2020), and code-switched sentences inducing cost effects at the processing level (Gross et al., 2019). Additionally, García and colleagues explored switch costs in noisy conditions, showing that highly proficient Spanish-English bilinguals tend to perform worse with mixed language conditions compared to single language conditions (García et al., 2018). While these studies investigated code-switching phenomena in human-to-human interactions, the role that different linguistic cues play in switch anticipation warrants further work to assess how listeners perform with voice-AI devices.

A handful of studies have begun to address related issues in this line of work. For example, Yu et al. (2023) outline and test TTS generation methods to improve Transformer-Transducer (T–T) models of Mandarin-English texts, against a lack of spoken data. Additionally, Hu et al. (2023) focus on improving ASR models through code-switching data from large language models. While these studies do not focus on intelligibility of generated code-switching, their methodological incorporation of TTS output of code-switching data motivates the present study to assess how listeners perform with digitally-generated speech. Addressing this facet of TTS voice development can be of use to related work looking to improve digitized speech output that is used for further purposes, including ASR.

In this realm of work, a key component to consider is the quality of TTS speech output and how that affects intelligibility among listeners. For example, Cohn and Zellou (2020) found that while listeners rated neural TTS methods as sounding more naturalistic, they performed better in concatenative TTS conditions. Such observations have important implications when it comes to the different speech generation methods that are used in different devices. Additionally, this adds further motivation to our work about what types of TTS methods might suit speakers and listeners who use more than one linguistic variety in the same conversational contexts. Our study brings together these remaining questions to consider how TTS voice quality conditions affect perception between monolingual and code-switched sentences.

The current study was designed to test the specific question of how code-switching in a TTS voice affects users' comprehension of the utterance, as well as how the TTS method might mediate perception of code-switched sentences. We played English-Spanish bilingual sentences generated using two different types of TTS methods (concatenative and neural) and asked them to transcribe the final word they heard. Across trials, sentences varied as being monolingual or code-switched and also whether the keywords were produced in English or Spanish. Consistent with prior work from human-human interactions, we predict that code-switched sentences should result in lower performance than monolingual sentences, due to effects stemming from potential lack of phonetic cues that are important in switch anticipation (Gross et al., 2019; Shen et al., 2020). Moreover, we predict that the TTS method should mediate this effect - the more difficult-to-understand TTS speech will result in even lower performance in code-switched sentences. Observing how code-switching and voice-AI quality affect user comprehension is important for both theoretical and practical reasons. From a theoretical perspective, it is important to understand whether the same types of language processing patterns from human-human interaction carry over into interactions with digital devices. Observations in this study and future studies may shed more light on how human listeners cognitively process different types of linguistic information from generated speech. Additionally, these findings may be helpful to researchers interested in generating large language models, which should ultimately be inclusive of fluid speech patterns as well. For example, differences between human-human and human-device interactions in switch contexts will warrant researchers in these areas to better understand how to model TTS code-switching to be reflective of human code-switching.

Practically, our study is informed by calls for better recognition of diverse speech patterns in voice-AI devices (Cihan et al., 2022; Zellou and Holliday, 2024) can be informative to the engineering and design of voice-enabled systems which permit code-switching, a common and natural linguistic behavior of bilingual language users. This not only serves as a type of representation for stigmatized and non-standardized linguistic practices, but can also be useful in translation and media transcription that rely on text-to-speech methods.

2 Materials and methods

2.1 Materials

To test these questions, we designed a word-identification task to assess how listeners perform with both monolingual and code-switched sentences for two TTS conditions, concatenative and neural. For the stimuli, 96 original English sentences were gathered from the speech-intelligibility-in-noise (SPIN) materials (Kalikow et al., 1977). Of these, 72 sentences were translated by the first author to create full Spanish stimuli and introduce intrasentential code-switches at utterance midpoints, yielding two linguistic modes (monolingual vs. code-switched) and two final-word targets (English vs. Spanish). This generated four conditions according to the final word and mode in the sentence, English, Spanish, English-Spanish, and Spanish-English. Sentence stimuli were then generated through Amazon's AWS TTS generator using a bilingual AWS Polly voice based on U. S. Spanish, Lupe, to create the switched and Spanish sentences. This resulted in 48 sentences being generated in two TTS conditions, concatenative and neural. In the end, this produced a full stimuli set of 12 sentences per target/mode condition in each TTS treatment. Additionally, following similar work examining intelligibility patterns and to increase the difficulty of the task, each stimulus was mixed with noise via Praat, with a sound-to-noise ratio at −3 dB, following similar procedures in prior work also examining speech intelligibility (Bradlow and Alexander, 2007; Clopper and Bradlow, 2008; Aoki et al., 2022).

2.2 Procedure and measurements

Participants were recruited through the University of California, Davis Psychology subject pool and received credit for their participation. 49 bilingual Spanish-English speakers (ages 19–30, 36 women, 11 men) completed a Qualtrics survey in which they listened to all 96 samples in randomized order. For each sentence, they were asked to type the final word. Listeners only heard the sentence once in order to control for potential effects of repetition aiding intelligibility. Additionally, participants completed a post-survey questionnaire to collect demographic information (e.g., gender, age, language use scores, etc.). Participants were specifically asked to rate their proficiency levels in a score out of 5 in speaking, listening, reading, and writing for both English and Spanish. Table 1 shows a full descriptive overview of these scores. Relevant to our study, 57.1% (*n* = 28) of participants reported a 5/5 rating for listening in Spanish, followed by 26.5% (*n* = 13) for 4/5 ratings, 12.2% (*n* = 6) for 3/5 and 4.1% (*n* = 2) for 2/5 ratings. No participant reported 1/5. For English listening proficiency scores, all participants reported either a 4 or 5 rating, with 89.8% (*n* = 44) reporting a 5/5 and 10.2% (*n* = 5) reporting a 4/5.

Participant responses were coded binomially for accuracy of word identification (1 = correct, 0 = incorrect). Data for the Spanish token *regazo* (lap) were excluded, due to investigator error causing a misspelling in token generation. All together, this yielded 4,679 trials

TABLE 1 Participants' language proficiency scores.

Domain	Spanish	English
Speaking	Average - 3.7 / 5	Average - 4.8 / 5
	Score percentages and counts	Score percentages and counts
	5–22.4% (*n* = 11)	5–83.7% (*n* = 41)
	4–44.9% (22)	4–16.3% (8)
	3–18.4% (9)	3–0
	2–10.2% (5)	2–0
	1–4.1% (2)	1–0
Listening	Average - 4.4 / 5	Average - 4.9 / 5
	Score percentages and counts	Score percentages and counts
	5–57.1% (*n* = 28)	5–89.8% (*n* = 44)
	4–26.5% (13)	4–10.2% (5)
	3–12.2% (6)	3–0
	2–4.1% (2)	2–0
	1–0	1–0
Reading	Average - 3.7 / 5	Average - 4.8 / 5
	Score percentages and counts	Score percentages and counts
	5–24.5% (*n* = 12)	5–85.7% (*n* = 42)
	4–34.7% (17)	4–12.2% (6)
	3–28.6% (14)	3–2.0% (1)
	2–8.2% (4)	2–0
	1–4.1% (2)	1–0
Writing	Average - 3.0 / 5	Average - 4.7 / 5
	Score percentages and counts	Score percentages and counts
	5–12.2% (*n* = 6)	5–75.5% (*n* = 37)
	4–14.3% (7)	4–20.4% (10)
	3–40.8% (20)	3–4.1% (2)
	2–22.4% (11)	2–0
	1–10.2% (5)	1–0

included for analysis. After data cleanup, an accuracy analysis was conducted through a linear mixed-effects regression model using the lmer package in R (Bates et al., 2015), to assess TTS and linguistic predictors. Fixed effects included TTS method (concatenative vs. neural), linguistic mode (switched vs. monolingual), and target word (Spanish vs. English), along with all possible interactions. Random effects included by-subject random slopes for TTS model and all linguistic conditions.

2.3 Predictions

Previous work has shown different rates of intelligibility for both concatenative and neural TTS voices (Cohn and Zellou, 2020), alongside other work showing switch costs in noisy conditions and in the absence of important phonetic cues (Shen et al., 2020). Informed by these findings, we predict that the concatenative condition will be more intelligible for listeners across all linguistic modes and target word conditions, and that code-switching conditions will induce lower intelligibility.

3 Results

Regression analysis (Table 2) shows that participants were overall accurate in their prediction of sentence-final target words (estimate = 0.90, pr < 0.001). For fixed effects, listeners performed worse in both the Spanish target condition (estimate = −0.76, pr < 0.001). They also performed worse in the code-switched contexts overall (estimate = −0.32; pr < 0.05), and in the code-switched contexts with Spanish target words (estimate = 0.35; pr = 0.05), though this interaction effect was marginal. No other fixed effects and interactions came back significant.

In addition to analyzing fixed and random effects for specific predictors, mean values of lexical identification and standard errors were analyzed as well. Figure 1 provides a visual comparison of mean values for TTS types across the two modes, separated by language of the lexical target, showing relatively better intelligibility in the English target word conditions.

Given participants' varied self-ratings on listening proficiency in Spanish, we also analyzed a subset of the data to observe any differences in identification accuracy for those who self-reported higher listening scores (score of 4 or 5 out of 5). 41 participants reported rating their listening skills in Spanish to be at least 4, excluding 8 participants. Regression analysis for the higher self-rated listeners showed only minimal difference in both regression analysis (Table 3) and mean difference visualizations (Figure 2). In this model, listeners still performed worse with the Spanish target words overall (estimate = −0.54; pr < 0.001), with no significant effects for code-switched contexts or interactions.

4 Discussion

Several key findings emerge in our study that raise important questions pertaining to past work and future directions. First, we predicted that there would be a statistically significant difference in intelligibility ratings between TTS methods. Specifically, we hypothesized that listeners would perform worse with neural TTS, and better with concatenative TTS. This is informed by Cohn and Zellou's (2020) study, in which listeners performed worse with neural TTS, even though they rated this method as more naturalistic-sounding. In contrast to their study, our findings show no significance for TTS across all conditions. There are a few possible reasons for this contrast, including population differences between each study, such as listeners' familiarity with both concatenative and neural TTS methods in voice-enabled devices. Additionally, when considering the lack of significance in the switch conditions, it could be the case that TTS methods incur less switch costs compared to human speech. We would expect this for the concatenative condition at the very least, due to less reduction in its TTS generation. Still, this does not explain why we do not observe significantly lower performance in the neural condition. In any case, future work should expand on this with a larger listener sample representing different ranges of voice-AI familiarity and linguistic experiences to better understand intelligibility of TTS. Additionally, comparative analyses should be conducted to look further at perceptual differences of code-switching between human-human and human-device interactions.

While TTS was not significant, our findings do show that listeners overall performed worse with Spanish target words, raising several questions about the nature of intelligibility with digital non-English voices. Alongside this, when considering previous work on switch costs in human-human interactions (García et al., 2018; Gross et al., 2019), we find that there was a small significant effect for code-switched

TABLE 2 Regression model output for all listeners.

	Estimate	Std. Error	*z*-value	Pr(>|*z*|)
Intercept	0.90	0.15	6.14	< 0.001 ***
Neural TTS	−0.15	0.13	−1.13	0.26
Spanish target	−0.76	0.13	−5.85	< 0.001 ***
Code-switched	−0.32	0.13	−2.43	< 0.05 *
Neural TTS × Spanish target	0.09	0.18	0.47	0.64
Neural TTS × code-switched	−0.23	0.18	−1.28	0.20
Spanish target × code-switched	0.35	0.18	1.93	0.05.
Neural TTS × Spanish target × code-switched	0.30	0.26	1.16	0.25

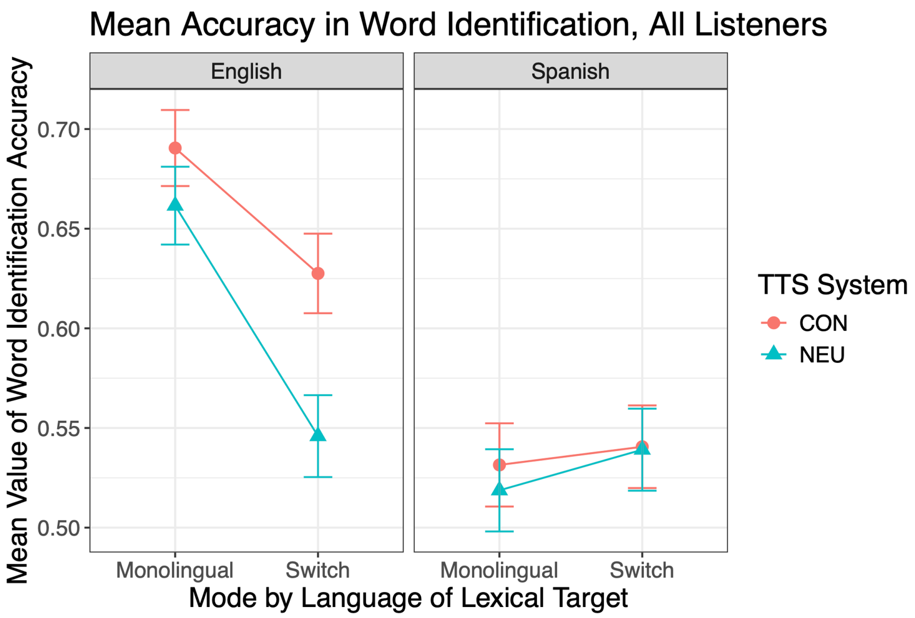

FIGURE 1
Mean values for correct word identification across all conditions for all listeners.

TABLE 3 Regression model results for subset of listeners who self-reported higher listening proficiency scores (4 or 5).

	Estimate	Std. Error	*z* value	Pr(>|*z*|)
Intercept	0.86	0.16	5.24	< 0.001 ***
Neural TTS	−0.16	0.15	−1.09	0.28
Spanish Target	−0.54	0.14	−3.75	< 0.001 ***
Code-switched	−0.23	0.14	−1.59	0.11
Neural TTS × Spanish target	0.07	0.20	0.33	0.74
Neural TTS × Code-switched	−0.25	0.20	−1.25	0.21
Spanish target × Code-switched	0.20	0.20	1.01	0.31
Neural target × Spanish target × Code-switched	0.31	0.28	1.10	0.27

conditions, confirming our prediction that intelligibility would be lower in those conditions due to potential switch costs in the perception of non-human speech. This motivates future work to continue assessing what phonetic cues are important for listeners in code-switching conditions, and to observe and compare switch-cost effect patterns between human-human and human-computer interactions.

Additionally, we collected information from each listener about their Spanish and English proficiency skills, with some listeners reporting higher listening scores in English. To observe any potential differences between listeners who self-reported lower vs. higher scores, we examined a subset of the data to assess only listeners who reported higher proficiency levels. Regression and mean accuracy analysis remained virtually unchanged; listeners performed worse still in the Spanish target word condition, with no significant effects for code-switching or TTS method. The reasons for overall lower intelligibility in Spanish target words, in both modes and TTS conditions, might be due to several reasons related to Spanish voice development in devices. For one, many of our listeners might be less familiar or have less daily frequency with Spanish digital voices, compared to English, influencing overall intelligibility over any effect the TTS methods might have had. This is not to say that the listeners in this dataset exhibited certain levels of linguistic ability, but rather that familiarity with digital Spanish voices might have an overall effect for all bilinguals, regardless of proficiency level. Additionally, it might be the case that the digital Spanish voice results in lower intelligibility due to being less developed compared to English digital

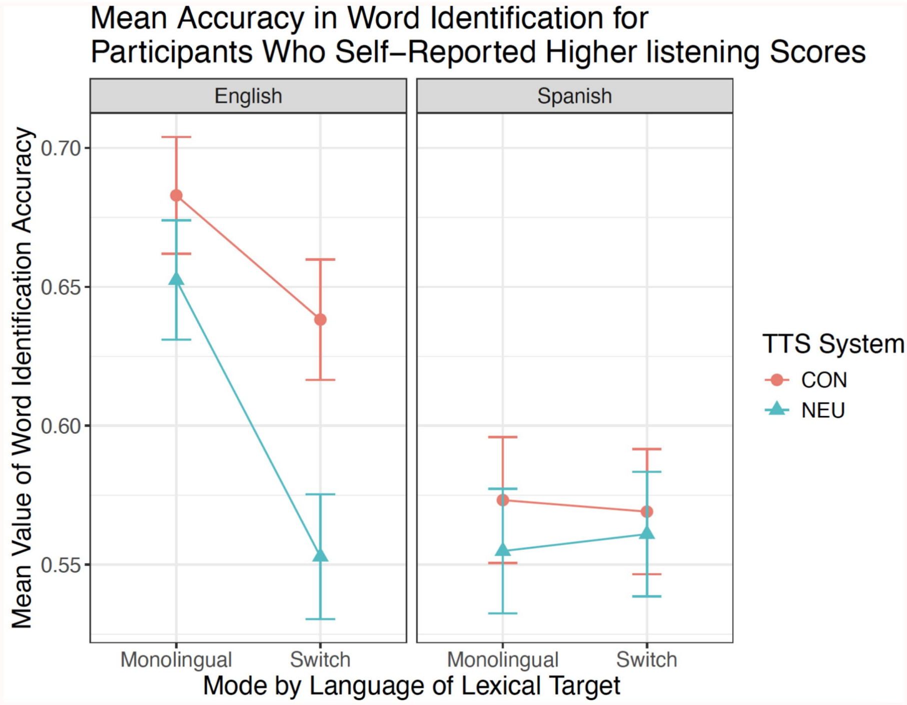

FIGURE 2
Mean values for subset of listeners who self-reported higher listening skills for Spanish (scores 4–5) for all conditions.

voices. One previous study has shown that transcription of different Spanish dialects in Amazon Alexa results in higher word error rates for U. S. Spanish, which is the variety used here (Nacimiento-García et al., 2024). While we expect that most of our listeners, being U.S. speakers, will have the most familiarity with the U.S. varieties of Spanish in human-human interactions, a lack of development on digital voices based on U.S. Spanish dialects could result in lower intelligibility overall. This encourages more work to see if this is indeed the case, and future studies should specifically look at different intelligibility patterns alongside word error rates for different Spanish accents. This could uncover several important points, including areas where enhanced TTS voice development is needed. Additionally, this advocates for more inclusivity within technological development of digital voices, to enhance intelligibility for human listeners and ASR across different varieties that represent natural linguistic diversity.

With these observations in mind, it is difficult to fully assess whether or not target word frequency and listener proficiencies affected intelligibility in the Spanish conditions. It might be the case that listeners' familiarity with voice-AI devices in Spanish affected their intelligibility with both TTS methods, or a combination of familiarity and word frequency of Spanish target words. In any case, our study highlights the ongoing need for more inclusivity of non-English varieties and fluid language practices in the design of voice-AI enabled devices. Outside of Spanish-English, code-switching is the norm for many speakers and is even represented in media, as in the case of Hinglish (Sailaja, 2011).

There were several limitations to this study that future studies could address. For one, the stimuli we used came from the SPIN test word list. Although this facilitated sentence generation, we did not control for word frequency, which may have an effect on bilingual listeners' processing of sentence-final words in noisy conditions. This, combined with switch costs from both TTS and noise, may have affected intelligibility rates, though this is difficult to assess from the data. Another key addition to future work is the inclusion of a set of human-voice trials in addition to the TTS stimuli. Incorporating a human comparison here would be beneficial to assess if the intelligibility differences we observe here are due to perceptual challenges or bias points in the TTS models. Additionally, we cannot be certain about each listeners' familiarity with each Spanish target word as bilinguals already exhibit a wide array of variation in linguistic production and perception. Future studies could both control for word frequency and assess how bilinguals perform with high and low frequency target words in their languages. Additionally, future work could generate stimuli in more than one SNR treatment to see if intelligibility varies in different noise conditions.

5 Conclusion

This study assessed intelligibility of code-switching in neural and concatenative TTS in noise. Our results show that Spanish target words were overall less intelligible than English target words across all conditions. Additionally, code-switched sentences were also less intelligible. The same findings emerged for the Spanish target words condition when observing the subset of listeners who reported higher listening proficiency. Additionally, the lack of significance in intelligibility rates between concatenative and neural TTS emerges in contrast to previous work (Cohn and Zellou, 2020) and motivates future work to continue investigating intelligibility of TTS across a wide range of linguistic contexts.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by UC Davis Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

TM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. GZ: Conceptualization, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aoki, N. B., Cohn, M., and Zellou, G. (2022). The clear speech intelligibility benefit for text-to-speech voices: effects of speaking style and visual guise. *JASA Express Lett.* 2:045204. doi: 10.1121/10.0010274

Auer, P. (1984). Bilingual Conversation. Amsterdam, Netherlands: John Benjamins.

Barker, G. (1947). Social functions of language in a Mexican-American community. *Acta Am.* 5, 185–202.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statistical Software* 67, 1–48. doi: 10.18637/jss.v067.i01

Bradlow, A. R., and Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J. Acoust. Soc. Am.* 121, 2339–2349. doi: 10.1121/1.2642103

Cihan, H., Wu, Y., Peña, P., Edwards, J., and Cowan, B. (2022). Bilingual by default: voice assistants and the role of code-switching in creating a bilingual user experience. In Proceedings of the 4th conference on conversational user interfaces (1–4). Association for Computing Machinery, New York, NY

Clopper, C. G., and Bradlow, A. R. (2008). Perception of dialect variation in noise: intelligibility and classification. *Lang. Speech* 51, 175–198. doi: 10.1177/0023830908098539

Cohn, M., and Zellou, G. (2020). Perception of concatenative vs. neural text-to-speech (TTS): differences in intelligibility in noise and language attitudes. In Proceedings of Interspeech. Interspeech Shanghai, China

Cooper, G. (2013). An exploration of intentions and perceptions of code-switching among bilingual Spanish/English speakers in the inland northwest. *J. Northwest Anthropol.* 47, 215–225.

García, P. B., Leibold, L., Buss, E., Calandruccio, L., and Rodriguez, B. (2018). Code-switching in highly proficient Spanish/English bilingual adults: impact on masked word recognition. *J. Speech Lang. Hear. Res.* 61, 2353–2363. doi: 10.1044/2018_JSLHR-H-17-0399

Gross, M. C., Lopez, E., Buac, M., and Kaushanskaya, M. (2019). Processing of code-switched sentences by bilingual children: cognitive and linguistic predictors. *Cogn. Dev.* 52:100821. doi: 10.1016/j.cogdev.2019.100821

Heller, M. (1988). Code-switching: anthropological and sociolinguistic perspectives. Berlin, Germany: Mouton de Gruyter.

Hu, K., Sainath, T. N., Li, B., Zhang, Y., Cheng, Y., Wang, T., et al. (2023). Improving multilingual and code-switching asr using large language model generated text. In 2023 IEEE automatic speech recognition and understanding workshop (ASRU) (pp. 1–7). IEEE. Taipei

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The J. Acoustical Society America* 61, 1337–1351. doi: 10.1121/1.381436

Kann, A. (2022). Voice assistants have a Plurilingualism problem. In Proceedings of the 4th conference on conversational user interfaces (1–5). Association for Computing Machinery. New York, NY

Kim, J., Merrill, K. Jr., Xu, K., and Kelly, S. (2022). Perceived credibility of an AI instructor in online education: the role of social presence and voice features. *Comput. Hum. Behav.* 136:107383. doi: 10.1016/j.chb.2022.107383

Mabule, D. R. (2015). What is this? Is it code switching, code mixing or language alternating. *J. Educ. Soc. Res.* 5, 339–350. doi: 10.5901/jesr.2015.v5n1p339

Nacimiento-García, E., Díaz-Kaas-Nielsen, H. S., and González-González, C. S. (2024). Gender and accent biases in AI-based tools for Spanish: a comparative study between Alexa and Whisper. *Appl. Sci.* 14:4734. doi: 10.3390/app14114734

Nilep, C. (2006). "Code switching" in sociocultural linguistics. *Colo. Res. Lingu.* 19:62. doi: 10.25810/hnq4-jv62

Piccinini, P. E., and Garellek, M. (2014) Prosodic cues to monolingual versus code-switching sentences in English and Spanish. In Proceedings of the 7th speech prosody conference (pp. 885–889) University of California, San Diego

Poplack, S. (2000). Toward a typology of code-switching. Wei, L. (ed.), The bilingualism reader. London, New York: Routledge, 221–255.

Sailaja, P. (2011). Hinglish: code-switching in Indian English. *ELT J.* 65, 473–480. doi: 10.1093/elt/ccr047

Shen, A., Gahl, S., and Johnson, K. (2020). Didn't hear that coming: effects of withholding phonetic cues to code-switching. *Biling. Lang. Congn.* 23, 1020–1031. doi: 10.1017/S1366728919000877

Toshniwal, S., Sainath, T. N., Weiss, R. J., Li, B., Moreno, P., Weinstein, E., et al. (2018). Multilingual speech recognition with a single end-to-end model. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4904–4908). IEEE. Calgary, AB

U.S. Census Bureau, U.S. Department of Commerce. (2023). Language spoken at home. American community survey, ACS 1-year estimates subject tables, table S1601. Available online at: https://data.census.gov/table/ACSST1Y2022.S1601?g=040XX00US01. (Accessed October 12, 2024)

Yu, H., Hu, Y., Qian, Y., Jin, M., Liu, L., Liu, S., et al. (2023). Code-switching text generation and injection in mandarin-english asr. In ICASSP 2023–2023 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 1–5). IEEE. Rhodes Island

Zellou, G., and Holliday, N. (2024). Linguistic analysis of human-computer interaction. *Front. Comp. Sci.* 6:1384252. doi: 10.3389/fcomp.2024.1384252

Zhan, X., Abdi, N., Seymour, W., and Such, J. (2024). Healthcare voice AI assistants: factors influencing trust and intention to use. *Proc. ACM Hum.-Comput. Interact.* 8, 1–37. doi: 10.1145/3637339