frontiers | Frontiers in Language Sciences

TYPE Methods
PUBLISHED 20 June 2025
DOI 10.3389/flang.2025.1569448

OPEN ACCESS

EDITED BY
Pedro Gomez-Vilda,
Neuromorphic Speech Processing Laboratory, Spain

REVIEWED BY
Agustín Álvarez-Marquina,
Polytechnic University of Madrid, Spain
Joshua Waxman,
Yeshiva University, United States
Wei Xue,
Saarland University, Germany

*CORRESPONDENCE
Robert L. MacDonald
✉ bmacdonald@google.com

RECEIVED 04 March 2025
ACCEPTED 16 May 2025
PUBLISHED 20 June 2025

CITATION
Martin A, MacDonald RL, Jiang P-P, Ladewig M, Cattiau J, Heywood R, Cave R, Tobin J, Nelson PC and Tomanek K (2025) Project Euphonia: advancing inclusive speech recognition through expanded data collection and evaluation. *Front. Lang. Sci.* 4:1569448. doi: 10.3389/flang.2025.1569448

Project Euphonia: advancing inclusive speech recognition through expanded data collection and evaluation

Alicia Martin[1], Robert L. MacDonald[1]*, Pan-Pan Jiang[1], Marilyn Ladewig[2], Julie Cattiau[1], Rus Heywood[1], Richard Cave[3], Jimmy Tobin[1], Philip C. Nelson[1] and Katrin Tomanek[1]

[1]Google Research, Mountain View, CA, United States, [2]CP Unlimited, New York, NY, United States, [3]Computer Science Department, University College London (UCL), London, United Kingdom

Speech recognition models, predominantly trained on standard speech, often exhibit lower accuracy for individuals with accents, dialects, or speech impairments. This disparity is particularly pronounced for economically or socially marginalized communities, including those with disabilities or diverse linguistic backgrounds. Project Euphonia, a Google initiative originally launched in English dedicated to improving Automatic Speech Recognition (ASR) of disordered speech, is expanding its data collection and evaluation efforts to include international languages like Spanish, Japanese, French and Hindi, in a continued effort to enhance inclusivity. This paper presents an overview of the extension of processes and methods used for English data collection to more languages and locales, progress on the collected data, and details about our model evaluation process, focusing on meaning preservation based on Generative AI.

KEYWORDS

disordered speech, automatic speech recognition, speech data collection, dysarthria, artificial intelligence

1 Introduction

Traditional ASR models, trained primarily on standard speech patterns, often fail to accurately interpret the diverse spectrum of human voices. This creates a communication barrier that can perpetuate social inequities, with serious consequences in healthcare settings where misinterpretations can lead to misdiagnosis, incorrect treatment, and even patient harm (Topaz et al., 2018). Project Euphonia, a Google Research initiative, is tackling this challenge by building the world's largest dataset of disordered speech. Using a proprietary web-based audio tool, Project Euphonia gathers speech samples from consented participants who record prompted phrases. As of February 2025, this dataset includes over 1.5 million utterances from ~3,000 speakers.

This paper outlines our approach to collecting and curating a high quality disordered speech dataset with the goal of supporting improved ASR accuracy for international languages and diverse speech patterns. Building on the success of our English-language dataset, we have expanded our efforts globally, capturing the rich diversity of speech within languages like Spanish, French, Japanese, and Hindi (Jiang, 2022). This global expansion allows us to create a high-quality, multilingual corpus for training more inclusive and accurate ASR models. Notably, prior research demonstrates that personalized ASR models

trained on Project Euphonia data can outperform human transcribers for individuals with disordered speech, highlighting the transformative potential of this approach (Green et al., 2021).

This work underscores the critical importance of incorporating disordered speech data into ASR model development. By building more inclusive datasets, we enable the creation of models that better serve users with diverse speech patterns, directly supporting initiatives like the Speech Accessibility Project, which aims to make voice-enabled technology accessible to all users, regardless of their speech characteristics (University of Illinois at Urbana-Champaign, 2024).

While Project Euphonia's data corpus is proprietary and accessible only to Google researchers for training and fine-tuning ASR models, this paper contributes to the broader community in three significant ways. Firstly, it details important considerations for curating datasets aimed at enhancing ASR across multiple languages. Secondly, it provides open-source resources to enable researchers and developers to replicate these data curation and ASR improvement processes. Thirdly, it introduces an alternative approach to analyzing ASR model performance in non-English languages, focusing on meaning preservation as a complement to the more traditional metric of word error rate.

2 Related work

Disordered speech encompasses a range of difficulties affecting intelligibility, fluency, articulation, and other aspects of vocal production (American Speech-Language-Hearing Association, 1993). Such difficulties can arise from various medical conditions, including Amyotrophic Lateral Sclerosis (ALS), Parkinson's Disease, and Down Syndrome, among others. The underlying causes often involve neurological conditions, developmental delays, or hearing impairments. The acute lack of representative speech datasets for these diverse and often highly individualized speech patterns further marginalizes these populations, limiting their ability to benefit from advancements in ASR technologies.

The curation of comprehensive disordered speech datasets for non-English languages is a recognized challenge, hindered by privacy concerns, collection costs, and a historical research focus on English (CLARIN, 2024). While valuable corpora exist, such as the multilingual AphasiaBank for aphasia research (including Spanish and French) and language-specific datasets like DysArinVox for Mandarin dysarthria (Zhang et al., 2024), PC-GITA for Spanish Parkinson's speech (La Quatra et al., 2024), or EasyCall for Italian dysarthric speech (Turrisi et al., 2021), a persistent gap remains. Specifically, there is a scarcity of large-scale datasets encompassing diverse etiologies across multiple non-English languages, which are crucial for effectively training robust Generative AI models.

Project Euphonia's recent efforts have begun to assess the performance of existing ASR and natural language understanding tools on newly gathered non-English disordered speech, for instance in Spanish and French, often by collecting utterances of pre-defined phrases. These evaluations provide important insights into current model capabilities and limitations on new linguistic contexts with disordered speech, with the intention of fine-tuning speech models specifically tailored to the nuances of these non-English languages for varied speech disorders. This paper builds upon this framework by providing a detailed look into the process of creating the dataset and evaluating performance on a small sample.

3 Expanding data collection

The goal of this phase of expansion is to understand enough about the data collection and curation processes in new languages and geographies to develop lessons learned that would enable much larger expansion over time. At the same time, the dataset specifications are designed to allow research on adapting multilingual ASR models to quantify key requirements for data quality and quantity of future data collections.

We expanded our data collection efforts to include speakers with disordered speech of four languages (Spanish, French, Japanese, and Hindi) from six countries (Mexico, Colombia, Peru, France, India, and Japan), to ensure a more inclusive representation of speech patterns across different languages and cultures. While some elements of the data collection were unchanged relative to collections of English speech, others required varying degrees of localization, including recruitment materials, consent forms, getting started guides, and prompt phrases.

Prior to recording, participants provided informed consent through our Project Euphonia web based audio tool. Consent forms were translated by Google Localization Services (Google, 2025). To ensure equitable compensation, participant payments were adjusted based on location and disbursed in local currency via Visa Gift Cards or PayPal, complying with US Office of Foreign Assets Control sanctions.

Our non-English dataset includes 38 Spanish speakers from Mexico, Colombia, and Peru; 14 French speakers; 76 Japanese speakers; and 4 Hindi speakers (Table 1). To reach a diverse range of speakers with disordered speech, we established strategic partnerships with organizations such as the International Alliance of ALS/MND, the Paris Brain Institute, and LSVT Global. These partnerships provided access to established networks of individuals with various health conditions affecting speech, facilitating recruitment and ensuring representation across multiple etiologies, speech impairments, and speech severities.

The observed imbalance in speaker distribution across target languages can be attributed primarily to our recruitment methodology. This approach relied on collaborations with various organizations supporting individuals with speech disorders. Partner organizations serving these populations drove data collections within a fixed timeframe as opposed to driving for a specific speaker or utterance count. This approach was used as a pilot for the localized data collection processes and tools. To achieve a more representative dataset as part of a scaled up program, a more targeted recruitment strategy would be essential, for example, by setting specific recruitment goals for underrepresented languages.

3.1 Phrase sets

Analogous to our previous data collections in English, this collection included exclusively scripted or read speech, collected through a web app that prompts participants with a phrase or

TABLE 1 Participant distribution by language (Overall $N = 132$).

Spanish	($N = 38$)
Etiology	
Amyotrophic lateral sclerosis	25 (65.8%)
Cerebral palsy	6 (15.8%)
Deaf	2 (5.3%)
Multiple system atrophy	1 (2.7%)
Stroke	1 (2.7%)
Unknown	3 (7.9%)
Speech disorder	
Dysarthria	30 (78.9%)
Deaf accented speech	2 (5.3%)
Dysphonia	1 (2.6%)
Within normal limits	5 (13.2%)
Speech severity	
Severe	13 (34.3%)
Mild	10 (26.4%)
Moderate	9 (23.7%)
Profound	1 (2.7%)
Within Normal Limits	5 (13.2%)
French	**($N = 14$)**
Etiology	
Parkinson's disease	1 (7.2%)
Amyotrophic lateral sclerosis	4 (28.6%)
Cerebral palsy	5 (35.8%)
Stutter	1 (7.2%)
Within normal limits	1 (7.2%)
Unknown	2 (14.3%)
Speech disorder	
Dysarthria	10 (71.5%)
Speech sound disorder	1 (7.2%)
Hyponasality	2 (14.3%)
Within normal limits	1 (7.2%)
Speech severity	
Severe	4 (28.6%)
Mild	3 (21.5%)
Moderate	5 (35.8%)
Within normal limits	2 (14.3%)
Japanese	**($N = 76$)**
Etiology	
Amyotrophic lateral sclerosis	12 (15.8%)
Cerebral palsy	4 (5.3%)
Parkinson's disease	35 (46.1%)
Down syndrome	1 (1.4%)

(Continued)

TABLE 1 (Continued)

Japanese	($N = 76$)
Vocal cord paralysis	3 (4.0%)
Cleft plate	2 (2.7%)
Undetermined	1 (1.4%)
Within normal limits	18 (23.7%)
Speech disorder	
Dysarthria	51 (67.2%)
Dysphonia	3 (4.0%)
Sound disorder	1 (1.4%)
Vocal cord paralysis	1 (1.4%)
Cleft palate	2 (2.7%)
Within normal limits	18 (23.7%)
Speech severity	
Severe	8 (10.6%)
Mild	37 (48.7%)
Moderate	12 (15.8%)
Profound	1 (1.4%)
Within normal limits	18 (23.7%)
Hindi	**($N = 4$)**
Etiology	
Stuttering	4 (100%)
Speech severity	
Mild	1 (25.0%)
Moderate	3 (75.0%)

Other information such as age, sex, race/ethnicity, and education level were not collected.

text prompt they are instructed to read verbatim. Many of the phrase prompts were complete sentences. For each language, we developed distinct phrase sets tailored to specific use cases: conversational phrases (typically used in communication between individuals), assistant phrases (relevant for home automation and voice-activated systems), and caregiver phrases (commands typically used between individuals and caregivers). These categories align with those used in our earlier data collections (MacDonald et al., 2021), ensuring consistency across domains.

Project Euphonia's ASR research methodology prioritizes the collection of scripted phrases over longer, more complex sentences for several key reasons. This approach is instrumental in ensuring high transcript conformity and significantly reducing the inherent difficulties and costs associated with the accurate transcription of disordered speech (MacDonald et al., 2021). The use of standardized phrases facilitates controlled and comparable studies across different speakers and ASR models. Furthermore, shorter, often domain-specific phrases (e.g., for home automation or caregiver interactions) make the data collection task more manageable for participants with speech impairments. This methodology also allows for careful curation to ensure comprehensive phonetic coverage essential for robust ASR training, even if individual phrases

do not always form complete grammatical sentences, and is well-suited for developing personalized models from limited per-speaker data.

From a pool of 3,598 possible phrases, 300 were randomly selected and presented to each speaker in their target language. The translation of these sentences was meticulously undertaken to respect cultural nuances (e.g., ensuring distinctions between Peninsular and Latin American Spanish were observed), while the core lexical items within the source phrases were kept consistent across languages for experimental control. Although our current random phrase selection did not isolate specific phonetic contexts, future research exploring such controlled contexts may reveal more pronounced differences in speakers with disordered speech, particularly in under-resourced languages.

Recognizing the importance of cultural relevance in speech data, we also collaborated with Google Localization Services to ensure accurate and culturally sensitive translations of the phrase sets. This included incorporating regional references to TV shows, brands, and celebrities familiar to each target audience. For instance, in the Latin American Spanish dataset, we translated the generic phrase "The best song of the year is 'Bad Guy'" with "La mejor canción del año es 'El Jefe,'" a song with strong local resonance.

3.2 Quality control processes

To enhance data quality, we implemented a standardized audio data collection protocol, adapted from our English data collection program. This involved providing localized quick guides for participants with detailed instructions on minimizing background noise, maintaining a natural speaking tone, avoiding audio clipping, and utilizing a personal device for optimal recording quality. This approach helped control for variability in recording conditions and ensured the collection of clean, usable audio samples across all languages.

To ensure accurate and comprehensive assessment of speech characteristics across languages, we implemented a rigorous, multi-faceted quality control (QC) process involving both automated systems and human reviewers. All collected data was subjected to our multi-stage QC protocol, the outcomes of which affirm the data's suitability for research. This protocol commenced with an automated QC pipeline, embedded directly within our proprietary administrative interface, that systematically assessed recordings against predefined technical standards, evaluating aspects such as audio clarity, completeness, and the absence of significant noise or interference. This integration enabled real-time feedback and programmatic flagging or exclusion of data failing these initial checks.

Subsequently, data that passed these automated assessments underwent meticulous human review. This involved a certified speech-language pathologist (SLP) conducting thorough speaker evaluations to confirm participant eligibility and the integrity of the speech sample according to study criteria. As part of the evaluation process, SLPs conducted a preliminary screening of the first 30 recordings from each participant to identify typical speech patterns and annotated instances of secondary speakers. Furthermore, meticulous transcription QC was conducted involving SLPs who could enlist additional support for challenging utterances. To further safeguard data quality and participant privacy, dedicated study coordinators reviewed the dataset to remove any incidental personally identifiable information (PII).

Following these preliminary assessments, SLPs with specialized training in assessing and grading speech abnormalities meticulously evaluated each speech sample. Evaluations accounted for variations in pronunciation, articulation, and fluency. SLPs then inferred potential etiologies, speech disorders, and speech severities often corroborated by participant self-reports. We applied the same grading scheme used in our English dataset consistently across all targeted languages. SLPs rated each sample using a five-point equal-interval scale: Typical (no impairment), Mild Impairment, Moderate Impairment, Severe Impairment, and Profound Impairment (Jiang et al., 2024). We expect that the uncertainty in these ratings is higher in languages where our SLPs were not fluent compared to our English-only dataset but that severity rating estimates were nonetheless worth including in the analysis.

For our international data collection effort, we contracted two SLPs who reviewed each transcript in order to form a consensus on the actual transcript. Consequently, this consensus-driven methodology did not include a formal evaluation of inter-rater or intra-rater variability for this study. Complementing this human review of transcripts, our automated QC system, designed to detect technical issues such as non-speech segments and amplitude problems, also proved highly effective. This was evidenced by the low number of technical issues subsequently identified by SLPs during their review of data that had already passed the automated QC process.

3.3 Audio tool

Project Euphonia uses a proprietary web-based audio tool for data collection, with an open-source version available on GitHub (https://github.com/google/project-euphonia-audiotool) to support replication and adaptation by researchers and developers (GitHub, 2025). This proprietary web-based audio tool was adapted to support multiple languages including Spanish, French, Hindi, and Japanese. This included UI changes to give clear guidance to users in multiple languages navigating the audio tool, translated informed consent forms, and translated getting started guides.

Designed with accessibility in mind, the tool accommodates users with motor, vision, or cognitive impairments. Participants can record phrases, skip phrases, return to re-record previous phrases, and listen to an automated voice articulate the target phrase before they speak, a feature particularly beneficial for users with low vision (Figure 1).

The recorded data is subsequently managed through a proprietary administrative interface used by Google researchers and SLPs. This interface enables reviewers to monitor recording progress and assign new phrases to users if issues are detected, such as excessive background noise or blank recordings resulting from audio capture failures. This early detection mechanism gives the Google research team the opportunity to collaborate with

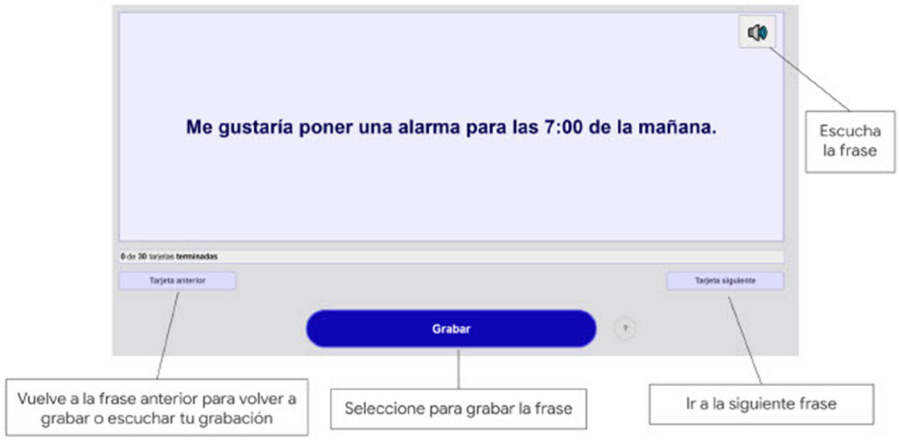

FIGURE 1
Project Euphonia web-based audio tool: displaying a phrase prompt in Latin American Spanish for participant recording.

participants to correct such issues, thereby ensuring higher data quality before the final dataset is curated for evaluating and training AI models.

4 Multilingual meaning preservation assessment using Gemini: a model evaluation process

This paper focuses on creating high-quality datasets rather than on developing ASR models. However, we did dedicate some time to comparing the output text from a personalized ASR model with that from an out-of-the-box ASR model (USM) for two international languages. In previous work, we have introduced an approach to automatically assess whether an ASR transcript captures the meaning of the original utterance based on a specifically fine-tuned version of a Large Language Model (LLM). LATTEScore, which builds on that, allows one to automatically assess the utility of an ASR model based on estimated meaning preservation (Tomanek et al., 2024).

Based on Project Euphonia's recent expansion in data collection across diverse linguistic landscapes as presented in this paper, we were able to extend our meaning preservation estimation approach to other languages as well. This allows for a more comprehensive evaluation of ASR models, including cutting-edge generative AI models like Gemini (Google, 2024), which are trained on a wide range of speech patterns. This improvement will make the system more accurate and accessible to a wider range of users, with particular benefits for individuals with disordered speech.

To evaluate the multilingual capabilities of our LATTEScore meaning preservation assessment approach, we created dedicated test sets for French and Spanish drawn from the Project Euphonia data corpus. These test sets comprised transcribed speech samples exhibiting a range of pronunciation variations, speech impairments, and other disordered speech patterns. A certified SLP reviewed the test set to ensure transcript accuracy. For Spanish, the test set consists of 518 examples from six speakers, while the French test set consists of 199 examples from 10 speakers. The disparity in speaker ratios between the two languages is due to differences in the severity of speech impairment across the groups, as we specifically excluded speakers with profound speech severity. Hindi and Japanese were not included in this evaluation, due to the unavailability of SLPs proficient in these languages to conduct the necessary human review of the ground truth transcriptions.

With the human-validated transcripts for our French and Spanish test sets established, we then generated transcripts from an out-of-the-box ASR model. This step was crucial because these ASR-generated transcripts served as input for our LATTEScore assessment, allowing us to evaluate how well our novel approach could identify and categorize meaning alterations present in typical machine-generated speech-to-text outputs.

To generate the automated transcriptions, we used Google's Universal Speech Model (USM; Zhang et al., 2023). Meaning preservation of this transcript relative to ground truth was quantified via a Primary Assessment Score, ranging from −1 to 2. This score indicated the severity of meaning alteration caused by transcription errors, with −1 signifying a perfect transcription and 2 highlighting serious inaccuracies that significantly impacted meaning (Figure 2).

Our human assessment of meaning preservation was based on scores from three non-SLP raters, each fluent in one of the target languages. For both Spanish and French, a human rated meaning preservation score was assigned. For Spanish, two fluent speakers collaborated to generate meaning preservation scores and primary error types. For French we employed a single rater. While this approach was practical given the dataset constraints, it may introduce bias and warrants re-evaluation in future research. While project constraints afforded only single ratings for each of these datasets, further research is warranted to estimate the potential bias or inter-rater variability.

Our meaning preservation classifier, based on the Gemini Nano-1 model (Google, 2023), achieved a high degree of accuracy (∼0.89 ROC AUC) on both French and Spanish test sets despite being trained solely on English examples (Table 2). This result highlights the robust multilingual capabilities of Gemini, enabling effective cross-lingual transfer without additional training

true_transcript	usm_ predicted _transcript	WER	primary_ assessment	primary_ error_type
Me gustaría poner una alarma para las 7:00 de la mañana.	mañana	0.91	2	deletion
Recuérdame que agarre un paraguas mañana por la mañana.	mañana ya	0.89	2	deletion
Apaga la tele.	apág	1.00	2	word error

FIGURE 2
Error analysis for Spanish data corpus.

TABLE 2 Model performance on language specific test sets (AUC-ROC).

Model	EN	ES	FR
Flan-cont-PaLM (62b)*	0.9	NA	NA
Gemini Nano-1	0.88	0.89	0.89

*Baseline model from previous research (Tomanek et al., 2024).

data (Tomanek and Martin, 2024). This capability opens up promising possibilities for future research and development, including more sophisticated cross-lingual applications beyond meaning preservation.

5 Future directions

Google is collaborating with the University of Illinois Urbana-Champaign (UIUC) along with four other tech companies on the Speech Accessibility Project (SAP), an initiative led by UIUC that is collecting and curating diverse datasets of disordered speech. These datasets are made publicly available to researchers and developers who sign UIUC's data use agreement and whose applications align with the program's objectives (University of Illinois at Urbana-Champaign, 2024). Project Euphonia contributes to this effort by providing advisory support on collecting disordered speech data, initially focusing on English and now expanding to Spanish, with the aim of encompassing a wide range of languages. This partnership highlights the importance of collaborative data sharing for research, ultimately enabling the development of more inclusive and effective speech recognition technology.

Furthermore, Project Euphonia is committed to fostering an open-source ecosystem that promotes broader data collection and enables the development of localized, personalized ASR models. As part of this commitment, researchers and developers can now access the Project Euphonia App, a suite of open-source resources designed to help create and customize speech recognition solutions for all languages. This toolkit provides software and documentation for key tasks such as personalizing open-source ASR models by fine-tuning, and deploying these models for transcribing speech. Crucially, the open-source tools, in their original form, are not intended for use without substantial modification for the diagnosis, treatment, mitigation, or prevention of any disease or medical condition. Developers bear sole responsibility for making such modifications and for ensuring that any applications they create comply with all applicable laws and regulations, including those pertaining to medical devices. To learn more visit: https://github.com/google/project-euphonia-app (GitHub, 2025).

This broader commitment to open science and community enablement is also evident in our support for initiatives like "tɛkyerɛma pa," a project with the University of Ghana and University College London's Center for Digital Language Inclusion (CDLI) that aims to enhance AI speech recognition for disordered speech in five major Ghanaian languages (Daily Guide Network, 2024). By fostering such collaborations, we promote inclusivity and accessibility, while expanding the linguistic diversity of data available for both research and product development.

6 Summary and limitations

The results of this study underscore the transformative potential of Project Euphonia's efforts to build diverse, high-quality datasets for disordered speech. By expanding data collection globally and collaborating with a wide array of organizations, we have contributed to improving the accuracy, inclusivity, and accessibility of ASR technologies for individuals with diverse speech patterns. These advancements have the potential to significantly improve the quality of life for individuals with speech disorders, enabling them to more easily interact with technology, access essential services, and participate in social, educational, and professional environments. Through better ASR, we move closer to creating a more equitable, inclusive world where communication is not a barrier.

6.1 Data collection tools

Project Euphonia primarily uses a proprietary web-based audio tool for data collection. A key contribution to the broader research community, however, is the release of an open-source version of this tool on GitHub (https://github.com/google/project-euphonia-audiotool) (GitHub, 2025). This resource enables researchers to replicate and standardize data collection efforts for disordered speech across new languages.

By leveraging this open-source toolkit, developers can streamline workflows, improve data quality, and adapt the

interface for multilingual use. However, while these tools offer efficiency and consistency, they may lack the flexibility required to fully reflect linguistic and cultural nuances, especially in low-resource settings. In such contexts, existing solutions may fall short, highlighting the importance of developing localized tools that can better accommodate regional speech patterns and user needs. This includes tailored interfaces, culturally appropriate prompts, and instructions in local languages. Ultimately, this trade-off calls for further research into when general-purpose tools suffice and when bespoke solutions are necessary to ensure both inclusivity and data accuracy.

6.2 Multilingual assessment

A key methodological contribution of this work is the successful extension and application of our meaning preservation assessment approach to new target languages (French and Spanish) using the Gemini Nano-1 model. The promising performance of the Gemini Nano-1 model in multilingual meaning preservation assessments highlights the future potential of ASR models in overcoming communication barriers and fostering greater social equity for people with disordered speech.

While these results are encouraging, this study represents a preliminary exploration of Project Euphonia's approach to multilingual meaning preservation assessment. The current evaluation of ASR models was limited to two languages (French and Spanish), which constrains the generalizability of our findings. To fully assess the potential and limitations of this methodology, future research should focus on evaluating ASR models across a broader range of languages, including both high- and low-resource languages. This expanded evaluation would offer deeper insights into the linguistic scalability of meaning preservation assessments and guide the development of more inclusive ASR systems.

Data availability statement

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding author.

Author contributions

AM: Data curation, Methodology, Project administration, Writing – original draft. RM: Conceptualization, Data curation, Funding acquisition, Supervision, Writing – original draft. P-PJ: Conceptualization, Data curation, Methodology, Project administration, Writing – original draft. ML: Data curation, Methodology, Writing – review & editing. JC: Conceptualization, Supervision, Writing – review & editing. RH: Conceptualization, Data curation, Methodology, Software, Writing – review & editing. RC: Data curation, Methodology, Writing – review & editing. JT: Data curation, Software, Writing – review & editing. PN: Funding acquisition, Supervision, Writing – review & editing. KT: Conceptualization, Methodology, Software, Writing – original draft.

Funding

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

Acknowledgments

We express our sincere gratitude to all the participants who generously shared their voice samples with Project Euphonia throughout the years. Your contributions are invaluable to our research and the development of more inclusive speech technology.

Conflict of interest

AM, RM, P-PJ, JC, RH, JT, PN and KT were employed at Google Research.

The remaining 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 author(s) declare that Gen AI was used in the creation of this manuscript. Gemini Advanced 1.5 Pro assisted with editing this manuscript.

Publisher's note

References

American Speech-Language-Hearing Association (1993). *Definitions of Communication Disorders and Variations*. Available online at: https://www.asha.org/policy/rp1993-00208/ (accessed January 31, 2025).

CLARIN (2024). *Corpora of Disordered Speech*. CLARIN RESOURCE Families. Available online at: https://www.clarin.eu/resource-families/corpora-disordered-speech (accessed January 31, 2025).

Daily Guide Network (2024). *Google, UG, GDI Hub Partner To Expand AI-Powered Speech Recognition. Daily Guide*. Available online at: https://dailyguidenetwork.com/google-ug-gdi-hub-partner-to-expand-ai-powered-speech-recognition/ (accessed January 21, 2025).

GitHub (2025). *Euphonia: Improving Speech Recognition for Non-Standard Speech*. Google. Available online at: https://github.com/google/project-euphonia-audiotool and https://github.com/google/project-euphonia-app (accessed January 21, 2025).

Google (2023). *Google Gemini AI: Advancing AI for all*. Google Blog.

Google (2024). *Introducing Gemini: The Future of AI, Powered by Google*. Google Blog.

Google (2025). *Localization and Internationalization*. Google Support. Available online at: https://support.google.com/l10n/?hl=en#topic=6307483 (accessed January 21, 2025).

Green, J. R., MacDonald, R. L., Jiang, P.-P., Cattiau, J., Heywood, R., Cave, R., et al. (2021). Automatic speech recognition of disordered speech: personalized models outperforming human listeners on short phrases. *ResearchGate*. 4778–4781. doi: 10.21437/Interspeech.2021-1384

Jiang, P.-P. (2022). *Project Euphonia: Automatic Speech Recognition Research Expands to Include New Languages, Including Spanish*. Google Research Blog Post.

Jiang, P.-P., Tobin, J., Tomanek, K., MacDonald, R. L., Seaver, K., Cave, R., et al. (2024). "Learnings from curating a trustworthy, well-annotated, and useful dataset of disordered English speech," in: *Proceedings of Interspeech 2024* (Kos: International Speech Communication Association). doi: 10.21437/Interspeech.2024-578

La Quatra, M., Turco, M. F., Svendsen, T., Salvi, G., Orozco-Arroyave, J. R., and Siniscalchi, S. M. (2024). Exploiting foundation models and speech enhancement for Parkinson's disease detection from speech in real-world operative conditions. *arXiv:2406.16128v1*. doi: 10.21437/Interspeech.2024-522

MacDonald, R. L., Jiang, P.-P., Cattiau, J., Heywood, R., Cave, R., Seaver, K., et al. (2021). "Disordered speech data collection: lessons learned at 1 million utterances from project euphonia," in *Proceedings of Interspeech 2021* (Brno: International Speech Communication Association). doi: 10.21437/Interspeech.2021-697

Tomanek, K., and Martin, A. (2024). *Assessing ASR Performance with Meaning Preservation*. Google Research Blog Post.

Tomanek, K., Tobin, J., Venugopalan, S., Cave, R., Seaver, K., Green, J. R., et al. (2024). "Large language models as a proxy for human evaluation in assessing the comprehensibility of disordered speech transcription," in *Proceedings of ICASSP 2024* (Seoul). doi: 10.1109/ICASSP48485.2024.10447177

Topaz, M., Schaffer, A., Lai, K. H., Korach, Z. T., Einbinder, J., and Zhou, L. (2018). Medical malpractice trends: errors in automated speech recognition. *J. Med. Syst.* 42, 153–154. doi: 10.1007/s10916-018-1011-9

Turrisi, R., Braccia, A., Emanuele, M., Giulietti, S., Pugliatti, M., Sensi, M., et al. (2021). EasyCall corpus: a dysarthric speech dataset. *arXiv:2104.02542*. doi: 10.21437/Interspeech.2021-549

University of Illinois at Urbana-Champaign (2024). *Speech Accessibility Project. Beckman Institute for Advanced Science and Technology*. Available online at: https://speechaccessibilityproject.beckman.illinois.edu (accessed August, 2024).

Zhang, H., Zhang, T., Liu, G., Fu, D., Hou, X., and Lv, Y. (2024). "DysArinVox: DYSphonia and DYSarthria mandARIN speech corpus," in: *Interspeech 2024* (Kos: International Speech Communication Association). doi: 10.21437/Interspeech.2024-1452

Zhang, Y., Han, W., Qin, J., Wang, Y., Bapna, A., Chen, Z., et al. (2023). Google USM: Scaling automatic speech recognition beyond 100 languages. *arXiv*. doi: 10.48550/arxiv.2303.01037