Nos_Brais-GL: A FAIR Galician TTS Corpus for Neural Speech Synthesis

Adina Ioana Vladu[1], **Antonio Moscoso Sánchez**[1,3], **Carmen Magariños**[1,2], **María Perez Lago**[1], **Elisa Fernández Rei**[1]
[1]Instituto da Lingua Galega, [2]Departamento de Electrónica e Computación
[3]Centro Singular de Investigación en Tecnoloxías Intelixentes
Universidade de Santiago de Compostela
{adina.vladu, antonio.moscoso.sanchez, mariadelcarmen.magarinos, mariaperez.lago, elisa.fernandez}@usc.gal

Abstract

This paper introduces *Nos_Brais-GL*, a new open-access high-quality Galician speech corpus designed for the development of neural Text-to-Speech (TTS) systems. *Nos_Brais-GL* contains approximately 18 hours of professionally recorded male speech and a carefully curated set of utterances selected to ensure linguistic variation and phonetic and prosodic richness. Beyond its immediate application in synthetic speech generation, *Nos_Brais-GL* exemplifies good practices in TTS corpus design for lesser-resourced languages, emphasizing methodological transparency, open licensing, and interoperability.

Keywords: TTS corpus, Galician, text-to-speech, TTS, language resources, minority languages, FAIR data

1. Introduction

In recent years, neural Text-to-Speech (TTS) systems have achieved remarkable improvements in naturalness, intelligibility, and expressiveness, largely driven by the availability of high-quality, phonetically rich, and prosodically varied speech corpora, coupled with crucial advancements in neural architectures –such as attention-based models and neural vocoders– capable of modeling the complex relationships within them (Tan et al., 2021). However, for languages with fewer resources, such as Galician, the scarcity of these corpora has long represented a significant barrier to developing robust and inclusive speech technologies (Ramírez Sánchez and García-Mateo, 2022).

To address this gap, *Proxecto Nós*[1] –an initiative led by the Universidade de Santiago de Compostela in collaboration with the Galician regional government– was launched with the objective of building a comprehensive ecosystem of linguistic resources for Galician (de Dios-Flores et al., 2022; Vladu et al., 2022). Following the FAIR data principles (Findable, Accessible, Interoperable, and Reusable) as presented in Wilkinson et al., 2016, the project seeks to ensure that all resources are openly available, reusable, and interoperable within broader European infrastructures for digital language equality.

This initiative is currently part of the national strategy to strengthen the digital presence of Spain's official languages. Alongside *Proxecto Nós* for Galician, similar efforts include *Projecte AINA*[2] for Catalan, *GAITU*[3] for Basque and *VIVES*[4] for Valencian. These programmes converge under *Project ILENIA (Impulso de las Lenguas en Inteligencia Artificial)*[5], which coordinates the creation of shared databases, infrastructures, and standards. Building on these foundations, the ongoing *Project ALIA*[6] aims to establish a public AI infrastructure with open and transparent language models, promoting the development and use of all official languages in Spain in global artificial intelligence ecosystems.

As a lesser-resourced language, Galician faces significant challenges in developing high-quality datasets due to the high economic and temporal costs involved. Although Galician benefits from a relatively solid foundation in textual resources, speech technologies remain underdeveloped, highlighting the need for open-access TTS corpora that preserve linguistic features such as pronunciation and prosody while promoting digital inclusion and cultural representation. Early initiatives in this field include the open-access Cotovía system (Rodríguez Banga et al., 2012) and the *CRPIH_UVigo-GL-Voices* dataset, comprising approximately 21 hours of multi-speaker recordings (CRPIH and GTM, 2023).

More recently, *Proxecto Nós* has advanced these efforts through the development of a large-scale high-quality single-speaker TTS corpus,

[1] https://nos.gal

[2] https://projecteaina.cat
[3] https://gaitu.eus
[4] https://vives.gplsi.es
[5] https://proyectoilenia.es
[6] https://alia.gob.es

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 9841–9849
11–16 May 2026.

Nos_Celtia-GL (Vázquez Abuín et al., 2023), and several neural TTS models, available through Hugging Face[7] and open to the public in the web interface *Nós-TTS*[8] (Magariños et al., 2024).

In this paper we introduce *Nos_Brais-GL*, a male single-speaker open speech corpus for high quality TTS in Galician. The resource is designed to bring methodological and linguistic improvements over the existing TTS resources, addressing some of their limitations. Specifically, *Nos_Brais-GL* improves upon the sentence modality balance, phonetic coverage, and linguistic naturalness of existing corpora, while maintaining high technical standards.

The corpus comprises approximately 18 hours of studio-quality speech, distributed across 16,121 utterances (168,000 words). It was recorded by a professional male voice talent selected through a rigorous perceptual evaluation involving 37 expert listeners who assessed clarity, articulation, prosody, and pleasantness. The textual material, compiled from diverse written and oral sources, was linguistically curated to maximize phonetic diversity, grammatical correctness, and representativeness of the contemporary usage of the Galician language.

The paper is organised as follows: Section 2 presents the corpus design and the methodology followed, Section 3 describes speech synthesis experiments carried out with state-of-the-art architectures, Section 4 highlights the results of the subjective and objective evaluation of the trained models' performance, Section 5 briefly presents the conditions under which the dataset is distributed and licensed, Section 6 offers concluding remarks, and Section 7 summarizes the ethical framework in which the work was carried out.

2. Corpus Design and Methodology

The *Nos_Brais-GL* corpus was carefully designed to ensure optimal quality in both text curation and voice selection. The text is phonetically rich, morphologically, lexically and prosodically varied, as well as thematically diverse, while the accompanying professionally selected male voice provides naturalness, clarity, and balance across the full range of Galician linguistic phenomena.

2.1. Speaker Selection

During the last few years, TTS systems have progressively been incorporated into everyday life through smart devices, playing an increasingly important role in daily tasks. On par, text-to-speech synthesis (TTS) technology has advanced significantly, improving speech fluency, intelligibility, and the ability to synthesize and convey emotions. The key challenge is that synthesized voices must not only sound natural, but also be pleasant to listen to, ensuring a satisfying user experience. For this purpose, the selection of a high-quality natural voice is a key step in the creation of synthetic voice models. Participants who contribute their voices to this process should meet specific linguistic and technical criteria, which include accurate pronunciation and intonation, sustained vocal endurance, and a stable speech rhythm (Braga et al., 2007a,b).

For the task of recording the *Nos_Brais-GL* corpus, the ideal speaker profile required a balance of professional experience, vocal stability, and linguistic competence in standard Galician. However, finding this speaker was not an easy task, due to the increasingly higher mistrust and uncertainty in the voice acting community regarding projects related to Artificial Intelligence. Due to this limitation, a non-professional voice was included in order to complete the group of four needed for the selection test. Finally, one speaker had to withdraw for personal reasons. The three remaining speakers were recorded in identical studio conditions, each reading a short selection of sentences representative of the corpus in terms of phonetic richness and prosodic variety.

Using the methodology described in García Díaz et al. (2024), a perceptual evaluation involving 37 linguists and speech-technology specialists was then conducted using the FOLERPA: Ferramenta On-Line para a ExpeRimentación PerceptivA[9] tool (Fernández Rei et al., 2021). Participants rated the recordings on five perceptual dimensions: (i) clarity and intelligibility, (ii) articulation accuracy, (iii) pleasantness, (iv) natural prosody, and (v) overall rating. The selected candidate (*Speaker 1*), achieved the highest ratings in the Pronunciation and Intonation categories, as well as the highest combined score. Figure 1 presents the results of the perceptual test.

Objective acoustic analysis confirmed the selected speaker's suitability for TTS purposes. The speaker maintained a stable rhythm of approximately 2.8 words per second and well-defined pauses, within the optimal range for synthetic voice modeling (Braga et al., 2007b,a), and a mean fundamental frequency of 172 Hz, the highest of the three candidates. From a phonetic standpoint, he consistently produced the seven Galician vowels, with clear mid-open/mid-close contrasts (/e/–/ɛ/, /o/–/ɔ/), and natural realizations of unstressed vowels. His consonant articulation was equally precise, notably in the production of /ʃ/ and /ɲ/, sounds that are often subject to dialectal or contact-induced variation (Regueira, 2009; González González, 2008).

[7] https://huggingface.co/collections/proxectonos/tts-models

[8] https://tts.nos.gal

[9] https://ilg.usc.gal/folerpa

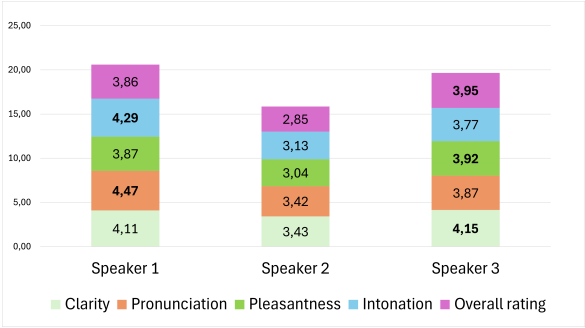

Figure 1: *Voice talents' scores for the five dimensions under evaluation: clarity, pronunciation, pleasantness, intonation, and overall rating*

Finally, in terms of prosody, the speaker naturally and consistently reflected the characteristic intonational patterns of Galician, particularly in the realization of interrogative utterances (Fernández Rei, 2016; Rodríguez Vázquez, 2019).

2.2. Text Corpus Compilation

The textual design of *Nos_Brais-GL* builds directly upon the methodology used in *Nos_Celtia-GL* (Vázquez Abuín et al., 2023) as presented in García Díaz et al. (2024), while addressing several of its limitations. The female corpus exhibited an imbalanced distribution of sentence modalities (especially lacking in exclamatory or imperative sentences), lexical repetition (particularly in numeric and formulaic phrases), and insufficient representation of specific phonetic contexts.

In *Nos_Brais-GL*, these aspects were systematically enhanced through corpus restructuring and targeted enrichment strategies. The resulting textual corpus comprises 16,121 sentences (168,320 words) organized into major subcorpora representing a range of textual domains (Table 1). The material brings together texts from institutional discourse (transcripts of sessions from the Galician Parliament - Parlamento de Galicia[10]), media and broadcasting (scripts from the region's main public broadcaster, Televisión de Galicia[11]), lexicographic sources (examples from the Real Academia Galega's dictionary[12])(González González, 2025), and transcribed oral materials from the Instituto da Lingua Galega's Arquivo do Galego Oral (AGO)[13](Fernández Rei, 2025), a reference archive for spoken Galician. It also incorporates content from a previously published TTS corpus developed by the Grupo de Tecnoloxías Multimedia / Centro Ramón Piñeiro para a Investigación en Humanidades (GTM/CRPIH)[14], which includes journalistic and linguistically constructed sentences designed to balance modality and prosody. The remaining utterances were selected from the same sources attending to length and modality criteria, forming a "Miscellaneous/Mixed" category and domain. Additionally, a dedicated subcorpus of 500 single words extracted from the Instituto da Lingua Galega's pronunciation dictionary, Dicionario de pronuncia da lingua galega (Regueira, 2025), was created to maximize phonetic coverage. Finally, approximately 1,500 utterances were extracted from literary works in order to enhance the representation of specific phonetic or prosodic contexts (e.g., alveolar realization of /n/, various types of imperative and exclamatory sentences).

Nevertheless, the *Nos_Brais-GL* textual corpus is significantly smaller than that of its female predecessor, since experiments with the *Nos_Celtia-GL* corpus showed that high-quality results could be obtained with as few as 13,000 sentences (Magariños et al., 2024). This optimization substantially reduced both the time and cost required for the production of this resource.

Finally, all texts were revised to conform to current Galician morphological and orthographic norms, correcting ungrammatical or unnatural constructions while ensuring fluency and linguistic consistency.

Source	Domain	Sentences
PD	Pronunciation dictionary	500
DRAG	Lexicographic	652
PG	Institutional	780
AGO	Oral	1,224
TVG	Media / broadcast	1,239
Misc	Literary	1,500
GTM	Miscellaneous / journalistic	4,151
Misc	Mixed	6,075

Table 1: Domain composition of the *Nos_Brais-GL* corpus. Source name acronyms correspond to: PD - Dicionario de pronuncia da lingua galega (ILG); DRAG - Dicionario da RAG; PG - Parlamento de Galicia; AGO - Arquivo do Galego Oral (ILG); TVG - Televisión de Galicia; GTM - GTM / CRPIH dataset; Misc - Miscellaneous.

2.3. Corpus Statistics and Analysis

This section presents a quantitative overview of the *Nos_Brais-GL* corpus, describing its sentence structure and phonetic balance to assess its suitability for TTS training.

[10] https://www.parlamentodegalicia.gal/Actividade/DiariosSesions

[11] https://crtvg.gal/gl

[12] https://academia.gal/dicionario

[13] http://ilg.usc.es/ago

[14] https://zenodo.org/records/8027725

2.3.1. Sentence Length and Modality

Sentence lengths were selected to approximate a Gaussian distribution, while also extending coverage at the short end and reducing excessively long, outlier sentences present in the feminine corpus. This design resulted in a distribution curve centered around 10–11 words per sentence, with a balanced representation of short (2–5 words), medium (6–15 words), and long (16–26 words) sentences, with the exception of a deliberate overrepresentation of one-word sentences to ensure correct pronunciation of very short utterances by subsequent TTS models.

The modality balance also shows improvement in *Nos_Brais-GL*, with an increase in interrogative sentences from 12.06% to 15.02% and, more significantly, in exclamatory utterances from 4.3% to 11.51% (Table 2).

Sentence type	Nos_Brais-GL (%)	Nos_Celtia-GL (%)
Declarative	67.07	80.28
Interrogative	15.02	12.06
Exclamative	11.51	4.3
Containing ellipsis	3.05	2.5
Mixed type	3.35	0.78

Table 2: Comparison of sentence modality distributions between the *Nos_Brais-GL* and *Nos_Celtia-GL* datasets.

2.3.2. Phoneme Statistics

The text corpus was phonetically transcribed with Cotovía[15] and dedicated scripts were used to analyze the number of phonemes and diphones, allowing for the detection and correction of distributional imbalances. To further ensure comprehensive diphone coverage, as explained earlier, a supplementary subcorpus of phonetically rich 500 single words was created using the CorpusCrt toolkit (Sesma and Moreno, 2000) and the Dicionario de pronuncia da lingua galega[16], guaranteeing a minimum of 30 instances per diphone (Kominek and Black, 2003).

The overall counts of phonemes and diphones (considering phrase-initial and phrase-final pauses and distinguishing between stressed and unstressed vowels), are presented in Table 3. Nearly all diphones in the corpus (99.75%) occur at least 30 times, while 98% of phonemes appear more than 5,000 times, confirming the corpus's phonetic richness both due to its size and to its design.

[15] https://gtm.uvigo.es/en/transfer/software/cotovia

[16] https://ilg.usc.es/pronuncia

Unit	Count	Total	Max a.fr.	Max n.fr.
Phonemes	38	823,261	64,076	77.83
Diphones	947	806,351	10,906	13.53

Table 3: Phoneme and diphone statistics in the *Nos_Brais-GL* corpus, presenting the number of different units, total counts, and maximum frequencies (absolute and normalized to 1,000).

2.4. Recording Protocol

All recordings were conducted at a professional recording studio under controlled acoustic conditions and standardized technical settings (48 kHz, 16-bit PCM WAV), using a single cardioid directional pattern microphone (NEUMANN U-87) to record the sound material, with the A/D plate set to 48kHz sampling frequency and 24-bit resolution. The workstation used by the recording technician was Pro Tools and the software used for subsequent normalization and editing was Izotope Rx. The duration of the recording sessions ranged from 2.5 to 5 hours, with breaks scheduled according to the needs of the speaker. Recording supervision followed detailed guidelines to ensure consistency. Linguistic prescriptions required the full reading of acronyms and numerals and the use of natural elisions and contractions. The speaker was instructed to articulate clearly and maintain a natural speaking style with a medium reading pace and moderate expressiveness. Interrogative and exclamatory sentences of the same type were read consistently to facilitate predictable prosodic modeling.

Each session was monitored in real time by trained linguists, who verified text-reading correspondence, phonetic accuracy, and prosodic coherence. All deviations were annotated, allowing for text adjustments where necessary, and retakes were performed when required to maintain homogeneity. The resulting recordings exhibit minimal noise, stable intensity and pitch levels, and consistent articulation and prosody across the 18-hour dataset.

Finally, three versions of the oral corpus were produced in WAV format: (1) a *raw* version containing the unedited recordings; (2) an *edited* version with mouth noises, clicks, and breaths removed, including 600 ms of initial and final silence; and (3) an *edited and normalized* version, scaled to a -10 dBTP level, prepared for direct use in model training.

3. Experiments and Evaluation

To validate the suitability of the *Nos_Brais-GL* corpus for TTS applications, we trained two distinct state-of-the-art models: VITS (Kim et al.) and Matcha-TTS (Mehta et al., 2024). These specific

architectures were chosen to benchmark the new dataset against both a widely established end-to-end system and a highly efficient recent approach.

VITS stands out as a fully end-to-end TTS framework that bypasses the traditional two-stage pipeline. It achieves this by jointly optimizing an acoustic model —based on conditional Variational Autoencoders (VAEs) (Kingma and Welling, 2014) and normalizing flows (Glow-TTS) (Kim et al., 2020)— alongside an adversarial vocoder (HiFi-GAN) (Kong et al., 2020). By combining Transformer-based encoders for robust linguistic feature extraction with normalizing flows to capture complex data distributions, VITS effectively generates high-fidelity audio. Furthermore, rather than relying on deterministic timing, the model employs a stochastic duration predictor, which introduces natural rhythmic variations and significantly improves the overall expressiveness of the synthesized speech.

Matcha-TTS (Mehta et al., 2024) employs a non-autoregressive encoder-decoder architecture for efficient acoustic modelling, trained via optimal-transport conditional flow matching (OT-CFM). This approach learns a deterministic Ordinary Differential Equation (ODE)-based generative process that rapidly maps a flow from Gaussian noise to target acoustic features, such as mel-spectrograms. By utilizing a text encoder, a duration predictor, and a U-Net-like decoder, the model efficiently generates high-quality acoustic features, avoiding the slow iterative processes inherent to traditional diffusion models. Finally, to synthesize the audio, this acoustic model is paired with the aforementioned HiFi-GAN vocoder, which converts the generated mel-spectrograms into high-fidelity waveforms, ensuring natural and highly realistic synthesized speech.

Both the VITS and Matcha-TTS models were trained on our corpus using grapheme-based representations. Text normalization for the grapheme inputs were handled by the Cotovía front-end (Rodríguez Banga et al., 2012). The VITS model underwent training for 1,300 epochs using the Adam optimiser with a learning rate of 0.0001 and a batch size of 48. Meanwhile, the Matcha-TTS model was trained following the authors' default hyperparameter configuration: 4,000 epochs, the Adam optimiser, a learning rate of 0.0001 and a batch size of 32.

To assess the naturalness and overall quality of the synthesized speech, we conducted an online Mean Opinion Score (MOS) test. We adopted the methodology and evaluation criteria previously established in García Díaz et al. (2024) to ensure consistency and comparability between *Nos_Brais-GL* and *Nos_Celtia-GL*. To evaluate model robustness, we selected a challenging test set of 21 sentences, all unseen during training. Following the approach taken in García Díaz et al. (2024), this selection was stratified to cover all sentence types (declarative, interrogative, exclamatory, elliptical), resulting in a subset of 14 declarative, four interrogative, one exclamatory, and one elliptical sentence. The set also included one highly complex example of spontaneous speech, designed to test a worst-case scenario: *Ai!, e xogando ó trompo tamén había xente que o collía á uña, non sabes?, e velo coller á uña...* (In English: *Oh! And when playing with the spinning top, some people would even catch it on their fingernail, you know? And seeing them do that...*).

The evaluation stimuli were generated by synthesizing the 21 test sentences with both models (VITS and Matcha-TTS), and the original recordings were included as the ground truth reference. To ensure the evaluation was concise, we adopted the sampling strategy from García Díaz et al. (2024). Each participant evaluated a total of 15 stimuli, which were composed of five randomly selected samples from each of the three conditions (original, VITS and Matcha-TTS). The presentation of these 15 stimuli was randomized for each listener.

We collected ratings on two five-point Likert scales, identical to those in García Díaz et al. (2024), assessing **quality** (from *1: Bad* to *5: Excellent*) and **naturalness** (from *1: Unnatural* to *5: Totally natural*).

A total of 34 participants took part in the test. The group consisted of 29 native Galician speakers and five high-proficiency speakers. Regarding technical familiarity, 20 participants identified as users of speech technology, while three identified as experts in the field.

To complement the subjective MOS evaluation, we performed an objective assessment using UTMOSv2 (Baba et al., 2024). Originally submitted as the T05 system for Track 1 of the VoiceMOS Challenge 2024 (Huang et al., 2024), this model was specifically designed to predict the naturalness MOS of high-quality synthetic speech. UTMOSv2 achieves this by combining spectrogram representations —obtained via a pre-trained image-based feature extractor— with acoustic features derived from self-supervised learning (SSL) speech models.

4. Results and Discussion

4.1. Subjective evaluation

Figure 2 shows the MOS results for the three evaluated conditions. As expected, the original human recordings (ground truth) received the highest scores, achieving 4.65 in quality and 4.37 in naturalness. These ground truth scores are notably higher than those obtained for the female corpus

in García Díaz et al. (2024) (4.37 and 4.10, respectively), suggesting either a higher perceived quality in the male voice recordings or a different listener expectation baseline.

Regarding the synthetic models, VITS performed strongly (Naturalness: 3.74, Quality: 3.99). Interestingly, this absolute performance is statistically identical in terms of naturalness, and comparable in quality, to the best model from García Díaz et al. (2024) (3.74 Naturalness / 4.19 Quality). This results in 'synthesis gaps' of 0.66 points (Quality) and 0.63 (Naturalness), which are substantially larger than the 0.18 and 0.36-point gaps for the female corpus, respectively. This difference is driven by the higher ground truth score in this work. This suggests that while VITS achieves a high absolute score, it struggled to capture the full quality of the male voice.

Within this study's models, VITS significantly outperformed Matcha-TTS (Naturalness: 3.38, Quality: 3.89) in terms of naturalness. The difference in quality, however, was not statistically significant at the 95% confidence level.

As observed in García Díaz et al. (2024), the overall synthetic scores were likely constrained by the inclusion of challenging sentence types (interrogative, exclamatory, and elliptical). The large synthesis gap for VITS, together with the lower performance of Matcha-TTS, suggests that both models may require further hyperparameter tuning specific to this corpus.

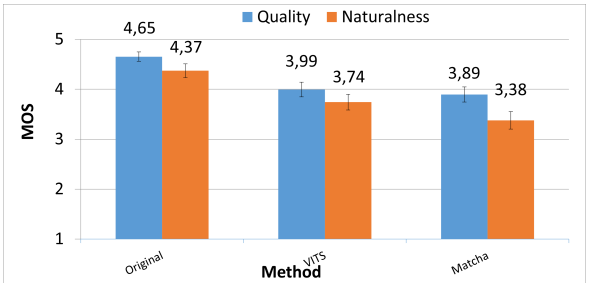

Figure 2: *MOS results for quality and naturalness, including 95% confidence intervals, for the original natural voice (Original), the VITS model (VITS), and the Matcha-TTS model (Matcha).*

4.2. Objective evaluation

As shown in Figure 3, Matcha-TTS achieved the highest UTMOSv2 score (3.55) among the synthesized systems, notably exceeding its subjective naturalness score of 3.38 (Figure 2). However, this objective ranking completely contradicts human perception, where listeners clearly preferred VITS over Matcha-TTS. Furthermore, UTMOSv2 severely underestimates both the natural voice baseline (scoring it 3.66 compared to the subjective 4.37) and the VITS model (3.22 vs. 3.74). We attribute this lack of correlation to a severe cross-lingual mismatch. Trained predominantly on English datasets —e.g., BVCC (Cooper and Yamagishi, 2021), SOMOS (Maniati et al., 2022)— UTMOSv2 is poorly calibrated for the phonetic and prosodic characteristics of our Galician corpus, rendering its objective predictions unreliable for this specific scenario.

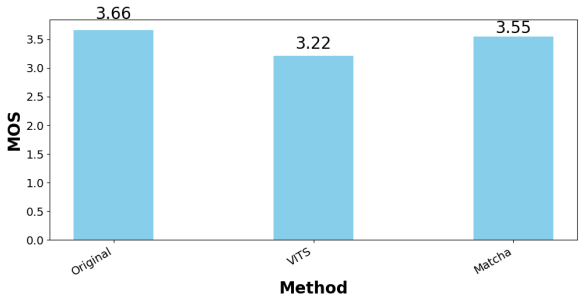

Figure 3: *Objective naturalness scores predicted by UTMOSv2. The bar chart compares the predicted MOS for the original human speech against the samples synthesized by VITS and Matcha-TTS.*

5. Availability and Licensing

The *Nos_Brais-GL* corpus (Vladu et al., 2025) is publicly available through the Zenodo research data repository[17] and the HuggingFace AI community[18]. The TTS models trained on this dataset are also available on Hugging Face[19] [20].

The dataset is released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, allowing unrestricted reuse, redistribution, and adaptation, provided that proper attribution is given to the creators. It was created according to the FAIR data principles (Wilkinson et al., 2016): it is findable via a permanent DOI and descriptive metadata on Zenodo; accessible in standard formats (WAV and CSV); interoperable through UTF-8 encoding, consistent metadata fields, and cross-reference with existing corpora such as *Nos_Celtia-GL*; and reusable thanks to detailed documentation and open license.

The distribution package on Zenodo includes:

- 18 hours of raw, edited, and normalized audio files in 48 kHz/24-bit WAV format;
- A metadata file in .csv format containing utterance identifiers and transcriptions;

[17] https://zenodo.org/records/14265241

[18] https://huggingface.co/datasets/proxectonos/Nos_Brais-GL

[19] https://huggingface.co/proxectonos/Nos_TTS-brais-matcha-graphemes

[20] https://huggingface.co/proxectonos/Nos_TTS-brais-vits-graphemes

- Documentation describing audio characteristics, file naming conventions, etc.

6. Conclusions

In this paper we have presented *Nos_Brais-GL*, a high-quality single-speaker Galician TTS dataset. *Nos_Brais-GL* represents a step forward in the creation of high-quality speech resources for Galician. Designed according to rigorous linguistic and technical criteria, the corpus provides a phonetically rich and prosodically diverse dataset that complements an existing resource, the *Nos_Celtia-GL* female corpus.

Beyond its immediate technical application, the corpus contributes to the digital sustainability of the Galician language and supports broader efforts toward linguistic equality in digital environments. By following FAIR data principles and open-licensing practices, it reinforces a reproducible and collaborative approach to resource development within the Iberian and the wider European language technology community.

7. Ethical and Legal Considerations

The *Nos_Brais-GL* corpus was developed in full compliance with ethical, legal, and data protection standards, in line with the FAIR principles and the open-licensing framework described in Section 5. All textual materials were incorporated under formal rights agreements with their respective owners and participating institutions, ensuring lawful reuse and redistribution.

All speakers involved in the recording process were duly compensated for their work and signed informed consent agreements transferring the necessary rights for the use and dissemination of their voice recordings.

Participants in the perceptual evaluation tests were informed of the study's purpose and their privacy rights prior to participation. All personal data were collected, processed, and stored in accordance with the European General Data Protection Regulation (GDPR, Regulation (EU) 2016/679), ensuring the protection of participants' anonymity and data integrity.

8. Acknowledgements

The authors wish to express their gratitude to Gaspar González Somoza for generously providing his voice for this project, to the speech talent who participated in the selection process, and to the listeners who took part in the perceptual tests.

Special thanks are due to the technical and linguistic team of the *Proxecto Nós* for their contribution to the dataset.

Our gratitude extends to the creators of the CorpusCrt tool and the Aholab Signal Processing Laboratory for making their resources available for phonetic balance analysis. We further acknowledge the institutions that contributed textual materials: *Grupo de Tecnoloxías Multimedia* (GTM), *Centro Ramón Piñeiro para a Investigación en Humanidades* (CRPIH), *Real Academia Galega*, *Corporación Radio Televisión de Galicia*, *Parlamento de Galicia*, and the ILG's *Arquivo do Galego Oral* and *Dicionario de pronuncia da lingua galega*.

This research was carried out within *Proxecto Nós*, funded by the Ministerio para la Transformación Digital y de la Función Pública and Plan de Recuperación, Transformación y Resiliencia - Funded by EU – NextGenerationEU within the framework of the projects ILENIA (ref. 2022/TL22/00215337) and Desarrollo Modelos ALIA.

The support of the Galician Ministry for Education, Universities and Professional Training and the "ERDF A way of making Europe" is also acknowledged through grant "Centro de investigación de Galicia accreditation 2024-2027 ED431G-2023/04."

9. Bibliographical References

Kaito Baba, Wataru Nakata, Yuki Saito, and Hiroshi Saruwatari. 2024. The t05 system for the voicemos challenge 2024: Transfer learning from deep image classifier to naturalness mos prediction of high-quality synthetic speech. *2024 IEEE Spoken Language Technology Workshop (SLT)*, pages 818–824.

Daniela Braga, Luís Coelho, Fernando Gil V. Resende, and Miguel Sales Dias. 2007a. Subjective and Objective Assessment of TTS Voice Font Quality. In *Proc. SPECOM 2007 – 12th International Conference Speech and Computer*, pages 306–311, Moscow, Russia.

Daniela Braga, Luís Coelho, Fernando Gil V. Resende, and Miguel Sales Dias. 2007b. Subjective and Objective Evaluation of Brazilian Portuguese TTS Voice Font Quality. In *Proc. AST 2007 - 14th International Workshop on Advances in Speech Technology*, pages 306–311, Maribor, Slovenia.

Erica Cooper and Junichi Yamagishi. 2021. How do voices from past speech synthesis challenges compare today? In *11th ISCA Speech Synthesis Workshop (SSW 11)*, pages 183–188.

CRPIH and GTM. 2023. CRPIH_UVigo-GL-Voices: Galician TTS dataset. https://doi.org/10.5281/zenodo.8027725. Dataset. Available under CC BY 4.0 license.

Iria de Dios-Flores, Carmen Magariños, Adina Ioana Vladu, John E. Ortega, José Ramom Pichel, Marcos García, Pablo Gamallo, Elisa Fernández Rei, Alberto Bugarín-Diz, Manuel González González, Senén Barro, and Xosé Luis Regueira. 2022. The Nós Project: Opening routes for the Galician language in the field of language technologies. In *Proc. of the Workshop Towards Digit. Lang. Equality within the 13th Lang. Resour. and Eval. Conf.*, pages 52–61, Marseille, France. ELRA.

Elisa Fernández Rei. 2016. Dialectal, Historical and Sociolinguistic Aspects of Galician Intonation. *Dialectologia*, 6:147–169.

Elisa Fernández Rei, Alba Aguete Cajiao, César Osorio Peláez, and Jose A. Cutrín Garabal. 2021. FOLERPA: Ferramenta On-Line para ExpeRimentación PerceptivA. https://ilg.usc.gal/folerpa.

Francisco Fernández Rei. 2025. Arquivo do galego oral. http://ilg.usc.es/ago. Director: Francisco Fernández Rei. Santiago de Compostela: Instituto da Lingua Galega. [Accessed: 2025].

Noelia García Díaz, Marta Vázquez Abuín, Carmen Magariños, Adina Ioana Vladu, Antonio Moscoso Sánchez, and Elisa Fernández Rei. 2024. Nos_Celtia-GL: An Open High-Quality Speech Synthesis Resource for Galician. In *IberSPEECH 2024*, pages 91–95, ISCA. ISCA.

Manuel González González. 2008. O novo galego urbano. In Mercedes Brea, Francisco Fernández Rei, and Xosé Luís Regueira, editors, *Cada palabra pesaba, cada palabra medía. Homenaxe a Antón Santamarina*, pages 363–374. Universidade de Santiago de Compostela, Santiago de Compostela.

Manuel González González. 2025. Dicionario da Real Academia Galega. https://academia.gal/dicionario. Director: M. González González. A Coruña: Real Academia Galega. [Accessed: 2025].

Wen-Chin Huang, Szu-Wei Fu, Erica Cooper, Ryandhimas E. Zezario, Tomoki Toda, Hsin-Min Wang, Junichi Yamagishi, and Yu Tsao. 2024. The voicemos challenge 2024: Beyond speech quality prediction. *2024 IEEE Spoken Language Technology Workshop (SLT)*, pages 803–810.

Jaehyeon Kim, Sungwon Kim, Jungil Kong, and Sungroh Yoon. 2020. Glow-TTS: A Generative Flow for Text-to-Speech via Monotonic Alignment Search. In *Proc. of the 34th Int. Conf. on Neural Inf. Process. Syst.*, NIPS'20, Vancouver, Canada. Curran Associates Inc.

Jaehyeon Kim, Jungil Kong, and Juhee Son. Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech. In *Int. Conf. on Machine Learning*.

Diederik P. Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations (ICLR)*, Banff, AB, Canada. Conference Track Proceedings.

J. Kominek and A. W. Black. 2003. The CMU ARCTIC Databases for Speech Synthesis Research. Technical report, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA.

Jungil Kong, Jaehyeon Kim, and Jaekyoung Bae. 2020. HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis. NIPS'20. Curran Associates Inc.

Carmen Magariños, Alp Öktem, Antonio Moscoso Sánchez, Marta Vázquez Abuín, Noelia García Díaz, Adina Ioana Vladu, Elisa Fernández Rei, and María Baqueiro Vidal. 2024. Nós-TTS: a Web User Interface for Galician Text-to-Speech. In *Proc. of the 16th Int. Conf. on Comput. Process. of Portuguese - Vol. 2*, page 200–203, Santiago de Compostela, Spain. ACL.

Georgia Maniati, Alexandra Vioni, Nikolaos Ellinas, Karolos Nikitaras, Konstantinos Klapsas, June Sig Sung, Gunu Jho, Aimilios Chalamandaris, and Pirros Tsiakoulis. 2022. SOMOS: The samsung open mos dataset for the evaluation of neural text-to-speech synthesis. In *Proc. Interspeech 2022*, pages 2388–2392.

Shivam Mehta, Ruibo Tu, Jonas Beskow, Éva Székely, and Gustav Eje Henter. 2024. Matcha-TTS: A fast TTS architecture with conditional flow matching. In *Proc. ICASSP*.

José Manuel Ramírez Sánchez and Carmen García-Mateo. 2022. Report on the Galician Language. Report D1.15, ELE.

Xosé Luís Regueira. 2009. Cambios fonéticos e fonolóxicos no galego contemporáneo. *Estudos de Lingüística Galega*, 1:147–167.

Xosé Luís Regueira. 2025. Dicionario de pronuncia da lingua galega. Director: Xosé Luís Regueira. Accessed 19/10/2025.

Rosalía Rodríguez Vázquez. 2019. La entonación de las preguntas parciales en una situación de contacto lingüístico: el caso del gallego y el español de galicia en hablantes bilingües. *Estudios de fonética experimental*, (28):81–124.

Eduardo Rodríguez Banga, Carmen García-Mateo, Francisco Méndez-Pazó, Manuel González-González, and Carmen Magariños. 2012. Cotovía: an open source TTS for Galician and Spanish. In *VII Jornadas en Tecnología del Habla and III Iberian SLTech Workshop, IberSPEECH*, pages 308–315.

A. Sesma and A. Moreno. 2000. CorpusCrt 1.0: Diseño de corpus orales equilibrados. http://gps-tsc.upc.es/veu/personal/sesma/CorpusCrt.php3. Computer program.

Xu Tan, Tao Qin, Frank K. Soong, and Tie-Yan Liu. 2021. A Survey on Neural Speech Synthesis. *ArXiv*, abs/2106.15561.

A. I. Vladu, N. García Díaz, X. L. Regueira Fernández, C. Magariños, A. Moscoso Sánchez, D. Fernández López, E. Fernández Rei, and F. Dubert-García. 2025. Nos_Brais-GL. https://doi.org/10.5281/zenodo.14265241. Dataset. Available under CC BY 4.0 license.

Adina Ioana Vladu, Iria de Dios-Flores, Carmen Magariños, John E. Ortega, José Ramom Pichel, Marcos Garcia, Pablo Gamallo, Elisa Fernández Rei, Alberto Bugarín, Manuel González González, Senén Barro, and Xosé Luis Regueira. 2022. Proxecto Nós: Artificial intelligence at the service of the Galician language. In *CEUR Workshop Proc.*, volume 3224, pages 26–30. CEUR-WS.

Marta Vázquez Abuín, Noelia García Díaz, Adina Ioana Vladu, Carmen Magariños, Adrián Vidal Miguéns, and Elisa Fernández Rei. 2023. Nos_Celtia-GL: Galician TTS corpus. https://doi.org/10.5281/zenodo.7716958. Dataset. Available under CC BY 4.0 license.

M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. da Silva Santos, P. Bourne, et al. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3:160018.