

X Congreso Internacional de Fonética Experimental

(Córdoba, 13-15 enero de 2026)

Sesión monográfica

Título	Voces sintéticas: nuevos desarrollos y retos para la fonética
Coordinación	Elisa Fernández Rei (Universidade de Santiago de Compostela)
Descripción	Justificación El desarrollo de voces sintéticas a través de arquitecturas neuronales ha experimentado un avance tecnológico notable en los últimos años, que ha traído consigo una mejora sustancial de la calidad y naturalidad de estas voces (Gupta et al. 2024, Ning et al. 2019)). Los sistemas basados en redes neuronales profundas han revolucionado el campo de la síntesis de voz, alcanzando niveles de realismo que en ocasiones resultan indistinguibles de la voz humana. Como es bien sabido, las tecnologías utilizadas para crear estos modelos de síntesis de voz son muy dependientes de los datos utilizados para su entrenamiento, primando con frecuencia la cantidad sobre la calidad y riqueza de dichos datos. Esta dependencia plantea interrogantes fundamentales sobre la representatividad lingüística, la diversidad dialectal y la preservación de las características fonéticas específicas de cada variedad lingüística. En este contexto, surge la necesidad de reflexionar sobre el papel que puede desempeñar la fonética en el desarrollo contemporáneo de las tecnologías del habla. ¿Tiene la fonética algo que aportar hoy en día al desarrollo de las tecnologías del habla y, en especial, de la síntesis de voz? Esta pregunta cobra especial relevancia cuando consideramos que los enfoques actuales, aunque exitosos desde el punto de vista tecnológico, pueden carecer de la fundamentación teórica que proporciona el conocimiento fonético profundo. Objetivos Esta sesión monográfica pretende abordar múltiples dimensiones de la intersección entre fonética y síntesis de voz: 1. Dar a conocer las innovaciones más importantes en el ámbito de la conversión texto-habla

	2. Proponer métodos innovadores para evaluar la naturalidad de las voces sintéticas actuales desde una perspectiva fonética rigurosa 3. Explorar propuestas teóricas y metodológicas desde la fonética que permitan mejorar: • La calidad de la síntesis mediante incorporación de conocimiento fonológico • La estabilidad de las realizaciones fonéticas • La naturalidad prosódica y temporal 4. Abordar metodologías para la creación de voces sintéticas que incorporen variación emocional, diferentes estilos de lectura y registros, adaptación a contextos comunicativos específicos y preservación de características dialectales y sociolingüísticas Perfil de participantes La sesión está dirigida a investigadores en fonética y fonología, especialistas en tecnologías del habla, desarrolladores de sistemas de síntesis de voz, profesionales de la industria tecnológica y estudiantes de posgrado en áreas afines.
Referencias	Chen, Z., & Min, X. (2023). Perceptual Quality Assessment of TTS-Synthesized Speech (pp. 423–435). https://doi.org/10.1007/978-981-99-0856-1_31 García Díaz, N., Vázquez Abuín, M., Magariños, C., Vladu, A.I., Moscoso Sánchez, A., Fernández Rei, E. (2024) Nos_Celtia-GL: an Open High-Quality Speech Synthesis Resource for Galician. Proc. IberSPEECH 2024, 91-95, doi: 10.21437/IberSPEECH.2024-19 Li, Z. K., Chen, M. M., Zhong, Y., Liu, P., & Duan, Z. (2024). GTR-Voice: Articulatory Phonetics Informed Controllable Expressive Speech Synthesis. 1775–1779. https://doi.org/10.21437/interspeech.2024-2216 Masson, Margot, and Julie Carson-Berndsen (2023). "Investigating phoneme similarity with artificially accented speech." In Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology, pp. 49-57. Oralbekova, D., Mamyrbayev, O., Kassymova, D., & Othman, M. (2024). Current advances and algorithmic solutions in speech generation. Vibroengineering PROCEDIA. https://doi.org/10.21595/vp.2024.23940 Qinsheng, D., Z. Jian, W. Lirong, and S. Lijuan (2011). "Articulatory speech synthesis: A survey", In Proc: 14th IEEE International Conference on Computational Science and Engineering, pp. 539–542. Reynolds, S., Machado, D., & Ochoa, O. (2024). A Transformer-Enabled Phonological Study of Genuine and Synthetic Voices. 19–26. https://doi.org/10.1109/aixset62544.2024.00009 Rodríguez-Banga, E., C. García-Mateo, F. Méndez-Pazó, M. González-González, and C. Magariños (2012). "Cotovia: an open source TTS for Galician and Spanish". In Proc. IberSPEECH 2012: VII Jornadas en Tecnología del Habla and III Iberian SLTech Workshop, pages 308–315. RTTH and SIG-IL.

	Sainz, Iñaki, Daniel Erro, Eva Navas, Inma Hernáez, Jon Sanchez, Ibon Saratxaga, and Igor Odriozola (2012). Versatile Speech Databases for High Quality Synthesis for Basque. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), pages 3308–3312, Istanbul, Turkey. European Language Resources Association (ELRA). Sainz, Iñaki, Ibon Saratxaga, Eva Navas, Inmaculada Hernáez, Jon Sanchez, Iker Luengo, and Igor Odriozola (2008). Subjective Evaluation of an Emotional Speech Database for Basque. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco. European Language Resources Association (ELRA). Saratxaga, Ibon, Eva Navas, Inmaculada Hernáez, and Iker Aholab (2006). Designing and Recording an Emotional Speech Database for Corpus Based Synthesis in Basque. In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy. European Language Resources Association (ELRA). Skerry-Ryan, RJ, Eric Battenberg, Ying Xiao, Yuxuan Wang, Daisy Stanton, Joel Shor, Ron J. Weiss, Rob Clark, Rif A. Saurous (2018). “Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron”, arXiv:1803.09047. https://doi.org/10.48550/arXiv.1803.09047 Streijl, R. C., S. Winkler, and D. S. Hands (2016). “Mean Opinion Score (MOS) revisited: methods and applications, limitations and alternatives,” Multimedia Systems, vol. 22, no. 2, pp. 213–227. Tan, X., T. Qin, F. Soong, T. Y. Liu (2021). “A Survey on Neural Speech Synthesis”, arXiv:2106.15561v3. Tokuda, K., H. Zen, and A. W. Black (2004). “An HMM-Based Approach to Multilingual Speech Synthesis”, In S. Narayanan and A. Alwan, editors, Text-to-Speech Synthesis: New Paradigms and Advances, chapter 7, pages 135–153. Prentice Hall. Zen, H., K. Tokuda, and A. W. Black (2009). “Statistical parametric speech synthesis”, Speech Communication, vol. 51, no. 11, pp. 1039–1154. Zhou, K., Zhao, S., Ma, Y., Zhang, C., Wang, H., Ng, D., Ni, C., Nguyen, T. H., Yip, J. Q., & Ma, B. (2024). Phonetic Enhanced Language Modeling for Text-to-Speech Synthesis. 3440–3444. https://doi.org/10.21437/interspeech.2024-359
Propuestas	Flores, Mariana, Eva Navas, Inma Hernáez, Iñigo Hierro & Asier Herranz: “Dialectal Personalization of TTS Systems: Experiments for Mexican Spanish”. Hierro Muga, Iñigo, Ibon Saratxaga Couceiro, Eva Navas Cordón, Inma Hernaez Rioja, Ander Arriandiaga, Laresgoiti & Victor Garcia Romillo: “Personalization strategies in Neural TTS: Evaluating pretrained model impact in Basque”. Magariños, Carmen & Antonio Moscoso Sánchez: “Voces sintéticas en gallego: estudio comparativo de distintas arquitecturas”. Vladu, Adina & Daniel Fernández López: “Evaluación de voces naturales para síntesis de voz en gallego”.