

1. Record Nr.	UNINA9911047827503321
Autore	Ekstein Kamil
Titolo	Text, Speech, and Dialogue : 28th International Conference, TSD 2025, Erlangen, Germany, August 25–28, 2025, Proceedings, Part I // edited by Kamil Ekštein, Miloslav Konopík, Ondej Pražák, František Pártl
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2026
ISBN	3-032-02548-6
Edizione	[1st ed. 2026.]
Descrizione fisica	1 online resource (654 pages)
Collana	Lecture Notes in Artificial Intelligence, , 2945-9141 ; ; 16029
Altri autori (Persone)	KonopikMiloslav PražákOndej PartlFrantisek
Disciplina	006.35
Soggetti	Natural language processing (Computer science) Natural Language Processing (NLP)
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	-- Speech. -- Lightweight Target-Speaker-Based Overlap Transcription for Practical Streaming ASR. -- An Empirical Analysis of Discrete Unit Representations in Speech Language Modeling Pre-training. -- Optimizing ASR Models with Semantic Information. -- Efficient Enhancement of Norwegian ASR Model. -- Towards Stable and Personalised Profiles for Lexical Alignment in Spoken Human-Agent Dialogue. -- Audio–Vision Contrastive Learning for Phonological Class Recognition. -- TOSD-Net: A CNN-Transformer Architecture for Robust Frame-Level Overlapping Speech Detection in Diverse Acoustic Conditions. -- An Exploration of ECAPA-TDNN and x-vector Speaker Representations in Zero-shot Multi-speaker TTS. -- Emotion-Aware Speech-Driven Facial Avatar Animation via Joint Blendshape Prediction and Emotion Recognition. -- Beyond Static Emotions: Leveraging Multitask Learning to Model Dynamics of Dimensional Affect in Speech. -- Implicit Speaker Group Encoding in Self-supervised Speech Recognition Models. -- Combining Temporal Visual Dynamics and Audio Representations for Robust Speaker Identification. -- Sentences vs Phrases in Neural Speech Synthesis: the Phrases Strike Back. -- Evaluating Phoneme-Level Pretraining in Czech Text-to-Speech Synthesis. -- Unifying Global and Near-Context Biasing in a Single Trie

Pass. -- Synthesising Cross-Speaker Data for Low-Resource Pathological Speech Recognition with PEFT. -- Multilingual Stutter Event Detection for English, German, and Mandarin Speech. -- How Far Can Synthetic Speech Go? Enhancing ASR in Low-Resource Scenarios via Voice Cloning. -- Enhancing Detection of Parkinson-induced Dysarthria with Cross-lingual Transfer Learning. -- Vocoder-Free Non-Parallel Conversion of Whispered Speech With Masked Cycle-Consistent Generative Adversarial Networks. -- Detection of Cognitive Disorders Using ASR-Based Nonsense Words Repetition. -- Mind the Gap: Entity-Preserved Context-Aware ASR for Structured Transcriptions. -- Boosting CTC-Based ASR Using LLM-Based Intermediate Loss Regularization. -- Robust Disfluency Labeling in Spontaneous Speech: Insights from Diverse Hungarian Corpora Including Mentally Ill Speakers. -- ParCzech4Speech: A New Speech Corpus Derived from Czech Parliamentary Data. -- Towards an Accurate Domain-Specific ASR: Transcription for Pathology. -- Automated Speaking Assessment for L2 Learners of Czech. -- Inclusive ASR for Critical Public Services: Debiasing with Actor-Simulated Speech. -- RECA-PD: A Robust Explainable Cross-Attention Method for Speech-based Parkinson's Disease Classification. -- Systematic FAIRness Assessment of Open Voice Biomarker Datasets for Mental Health and Neurodegenerative Diseases. -- When Silence Speaks: Understanding Open-Ended Responses via LLMs in Therapeutic Voice Interaction. -- Multilingual Domain Adaptation for Speech Recognition Using LLMs. -- Using Cross-attention For Conversational ASR Over The Telephone.

Sommario/riassunto

This conference volume constitutes the proceedings of the 28th International Conference on Engineering on Text, Speech, and Dialogue, TSD 2025, held in Erlangen, Germany, in August 2025. The 60 full papers were presented in this volume were carefully reviewed and selected from 122 submissions. They focus on speech and language technologies and computer processing of speech and language related data.
