8th International Conference on Artificial Intelligence and Speech Technology
AIST-2026
26th – 27th November, 2026

List of Registered Papers

Track 1: Speech Intelligence and Voice Generation Systems		
Sno	**Paper ID**	**Paper Title**
1	51	Audio Splicing Forgery Detection: A Systematic Review of Detection Methods, Datasets, and Trends
2	53	Transfer Learning-Based Speaker Identification for Low-Resource Sylheti Speech Using ECAPA-TDNN Embeddings
3	84	PrivSpeak: Deep Learning-based CNN architecture for Privacy-Preserving Speaker Recognition
4	185	TeleAPA: A Robustness Benchmark for Automatic Pronunciation Assessment under Teletherapy Transmission Conditions
5	192	Explainable AI for Speaker De-identification: Utility Preservation Without Privacy Gains
6	202	Deepfake voice detection with Android application
7	223	Closing the Indic ASR Gap
8	228	Cross-Corpus Dysarthria Detection: A Calibration Perspective on Self-Supervised Speech Representations
9	233	Comparative Analysis of Acoustic Feature Representations and Classical Machine Learning Algorithms for Bird Species Recognition
10	579	Empirical Evaluation of Output Representations and LM Fusion for Low-Resource Sanskrit ASR
11	581	A Multi-Model Deep Learning Framework for Deepfake Video Detection
12	616	Noise-Robust Speech Command Recognition Using MFCC Features and Machine Learning Classifiers
13	635	Dysarthric Speech Enhancement by Phoneme Replacement
14	637	Speaker and Corpus Independent Speech Emotion Recognition: An Unbiased Evaluation
15	642	TriVaani: Self-Transcript Tuning for Indic Speech Translation
16	719	The Carbon Cost of Speech Recognition: Quantisation Does Not Help, Location Does
17	729	Cross Corpus Speech Emotion Recognition Using Explainable AI and MAML
18	740	Cross-Modal Attention Fusion of Pretrained Speech and Facial-Expression Embeddings for Speaker-Independent Emotion Recognition on RAVDESS

19	741	Toward Text-to-Speech for Dhundhari: A Survey and Pivot-Based Morphological Adaptation Framework for a High-Density, Zero-Resource Rajasthani Dialect
20	806	An Accessible Bus Reservation System for Visually Impaired and Autistic People
21	822	DangauraSpeech: A Speech Corpus and Proximal Transfer Benchmark for Dangaura Tharu Automatic Speech Recognition
22	848	Explicit Grapheme-to-Phoneme Conversion for Hindi VITS: A Controlled Study of Phonemic and Graphemic Input
23	879	Acoustic-Based Alzheimer's Disease Detection from Spontaneous Speech: An Evaluation on the ADReSS 2020 Dataset
24	944	Confidence-Calibrated Risk-Aware ASR for Noisy Hindi--English Code-Switched Speech
Track 2: Multilingual and Human-Centric AI Interaction Systems		
Sno	**Paper ID**	**Paper Title**
25	27	DAWR: Dual-Stream Audio Word Recognition for Assamese Using MFCC-BiLSTM and Tiny Swin Transformer Fusion
26	49	KnightSpeak: An AI-Powered Multilingual Audio Chess Commentary System
27	262	Building a Multi-Domain Spoken Hindi Dataset for Automatic Text Summarization
28	585	Explainability for Multilingual Failure in Audio Deepfake Detection
29	587	BankAssist: A Multilingual Retrieval-Augmented Voice Assistant for Indian Banking Information Services
30	596	A Multilingual Speech-to-Indian Sign Language Translation System Using Fine-Tuned mT5 with Real-Time 3D Avatar Rendering
31	600	From Learners to AI-Supported English Teachers: Exploring English Communication Development through Multilingual AI Teacher Support (MATS) Model
32	774	ParkinsonXAI: An Explainable Voice-Based Machine Learning Framework for Parkinson's Disease Assessment and the Effect of Subject-Disjoint Evaluation
33	792	Speech-Based Parkinson's Disease Detection: A Comparative Study of Feature Selection, Oversampling and Machine Learning Classifiers
34	907	When Speaker Errors Become Attribution Errors: Measuring and Predicting the Downstream Cost of Diarization Failures in Multilingual Indian Speech
35	946	Breaking Language Barriers: Enhancing English-Hindi Translation with Transformer Models
36	954	Ṣaḍaṅga: A Six-Dimensional Interpretable Metric for Real-Time Assessment

		of Recited Speech
Track 3: Generative and Learning-Based AI for Speech Technologies		
Sno	**Paper ID**	**Paper Title**
37	479	FedAgent-AVSR: Agent-Guided Adaptive Fusion for Federated Audio-Visual Speech Recognition
38	554	LLM-Assisted Speech-Aware Reversible Data Hiding in Encrypted Images
39	567	Bridging the Dialects Gap: A Cross-Paradigm Benchmark of Multi-Task Learning Against Zero-Shot Arabic LLMs for Hate Speech Detection
40	615	VoiceRAG-GOV: A Governance-Aware Framework for Explainable and Trustworthy Speech-Based Enterprise Decision Intelligence
41	626	Cross-Lingual and Cross-Method Generalization Gaps in Audio Deepfake Detection: A Systematic Study on MLAAD-tiny
42	737	Generative AI and Speech Technology in Sustainable Healthcare Management: Policy Implications and Fraud Detection for Assisted Living
Track 4: Multimodal, Resource-Aware and Applied Intelligence		
Sno	**Paper ID**	**Paper Title**
43	169	Emotion-Aware Adaptive Learning: Multimodal Emotion Recognition and Psycho-Affective Feedback Generation
44	449	Hierarchical Two-Stage Framework for Separating Genuine Speech from Human Mimicry and AI Synthesised in Hindi
45	466	Artificial Intelligence (AI)- and Natural Language Processing (NLP)-Based Toxicity and Hate-Speech Detection and Moderation: A Review of Recent Advances
46	618	BhashaSetu: An AI-Based Multimodal Framework for Multilingual Speech and Video Localization in Indian Languages
47	624	Multimodal Speech Emotion Recognition with a PCA-Based Ensemble Classifier and a Matched Speaker-Leakage Analysis
48	634	Uncalibrated Fusion, Overstated Confidence: An Evaluation-Practice Survey of Multimodal Voice Phishing Detection
49	653	Subject-Independent EEG Classification of Om Chant Listening via Wavelet Features and Channel Selection
50	788	PayPulse: An Inclusive Multi-Modal Digital Wallet for Universal Banking and Accessibility, Empowering Visually Impaired and All Users
51	803	Multimodal Emotion Detection System Using Face, Voice, and Text Analysis
52	966	When Does the Evidence Decide? Evidence-Equivalent Scenarios for Measuring Unsupported Certainty in Multimodal Reconstruction of Missing Speech