

2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2025)

**Singapore
22-24 October 2025**

Pages 1-652



**IEEE Catalog Number: CFP2514U-POD
ISBN: 979-8-3315-7207-5**

**Copyright © 2025 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP2514U-POD
ISBN (Print-On-Demand):	979-8-3315-7207-5
ISBN (Online):	979-8-3315-7206-8
ISSN:	2640-009X

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

Table of Contents

Message from the Conference Chairs	XLV
Message from the Program Chairs	XLVII
Conference Committees	XLIX
Keynote Speeches	LXV
Tutorials	LXXI
Perspective Panel Sessions	LXXIX
Special Sessions	XCIII
Grand Challenge	XCIV
ANC Workshop	XCIX
3MT Competition	CII
Sponsors	CV
Author Index	CVII

Papers

Advanced Signal Processing and Machine Learning for Acoustic Scene Analysis and Signal Enhancement

D1-1430_L1.1 MVDR beamforming for underdetermined sound source separation using iterative PSD estimation in beamspace	1
<i>Jin Xuan Teh, Yusuke Hioka</i>	
D1-1430_L1.2 Exploring Dual-Mode Training for Real-time Target Speaker Extraction	7
<i>Li Li, Shogo Seki</i>	
D1-1430_L1.3 Switching Constant Separating Vector for Moving Source Extraction with Geometric Constraints	13
<i>Changda Chen, Yichen Yang, Yuehao Zhao, Shoji Makino, Jingdong Chen</i>	
D1-1430_L1.4 Neural Network-Assisted Joint DOA Estimation and Beamforming with First-Order Reflection Modeling	19
<i>Yichen Yang, Chao Pan, Qiang Gao, Jacob Benesty, Shoji Makino, Jingdong Chen</i>	
D1-1430_L1.5 Speaker Localization in Classroom Environments Using GCC-PHAT Features and Mamba State Space Models with Ad-hoc Microphone Arrays	24
<i>Rashed Iqbal, Christian Ritz, Jack Yang, Sarah Howard</i>	

D1-1430_L1.6 Joint Separation and Tracking of Moving Sources With Distributed Microphone Arrays Based on Time-Varying Inertial Spatial Models	30
<i>Ryunosuke Nihei, Yoshiaki Bando, Aditya Arie Nugraha, Diego Di Carlo, Hiroyuki Ueda, Yosuke Ito, Kazuyoshi Yoshii</i>	
D1-1430_L1.7 Visually-Informed Multichannel Sound Source Separation Based on 3D Gaussian Primitives	36
<i>Haruki Asano, Ryunosuke Nihei, Yoshiaki Bando, Aditya Arie Nugraha, Diego Di Carlo, Hiroyuki Ueda, Yosuke Ito, Kazuyoshi Yoshii</i>	
D1-1430_L1.8 Joint Optimization of Sampling Rate Offsets and Demixing Filters Using Auxiliary Function Method	42
<i>Hayato Takeuchi, Takao Kawamura, Nobutaka Ono, Shoko Araki</i>	
D1-1430_L1.9 First Demonstration of Acoustic Scene Classification Based on Trained Sound-to-Light Conversion	48
<i>Shun Kotsugi, Takao Kawamura, Nobutaka Ono</i>	
D1-1430_L1.10 Auxiliary-Function-Based Decentralized Independent Vector Analysis for Distributed Microphone Arrays	54
<i>Kouei Yamaoka, Katsuhiko Morita, Norihiro Takamune, Hiroshi Saruwatari</i>	
D1-1430_L1.11 Interactive Spatial Audio Rendering on Mobile Devices: A Two-Stage User Interface with Adaptive HRTF Selection and Real-Time Room Acoustics Simulation	60
<i>Shravan Raghunath, Kanishk AL, Sailesh S, Rishabh Gupta, Saurav Gupta, Ramesh R</i>	
D1-1430_L1.12 Are Identical Sounds Present in Distributed Recordings to Serve as Spatio-Temporal Anchors? A Case Study Using the SINS Database	66
<i>Takao Kawamura, Nobutaka Ono</i>	
Advanced Topics in Audio Understanding of Sound Events, Scenes, and Beyond	
D2-0830_L1.1 Evaluation of auditory and tactile perception for augmented sound-image enhancement using pre-virtual-leading hypersonic signals	72
<i>Ryota Imanaka, Yuting Geng, Masato Nakayama, Takanobu Nishiura</i>	
D2-0830_L1.2 Improvement in Variance Estimation in Variable-Step-Size Shared-Error NLMS Algorithm for Acoustic Echo and Noise Canceller	78
<i>Kenta Iwai</i>	
D2-0830_L1.3 Hierarchical Sparse Sound Field Reconstruction with Spherical and Linear Microphone Arrays	83
<i>Shunxi Xu, Craig T. Jin</i>	
D2-0830_L1.4 Robust Superdirective Beamforming Using a Uniform Circular Array with Directional Microphones	89
<i>Weilong Huang, Longfei Felix Yan, Emanuël A.P. Habets</i>	

D2-0830_L1.5 Towards Robust Stereo 3-D SELD: A Study of Perceptual Features and Data Augmentation	95
<i>Jun Wei Yeow, Ee-Leng Tan, Santi Peksi, Woon-Seng Gan, Huang Qirui</i>	
D2-0830_L1.6 Pre-training Autoencoder for Acoustic Event Classification via Blinky	101
<i>Xiaoyang Liu, Yuma Kinoshita</i>	
D2-0830_L1.7 Sound source enhancement using power spectral density estimation in beamspace for a dual unmanned aerial vehicle system	107
<i>Mingxue Song, Jin Xuan Teh, Yusuke Hioka, Benjamin Yen, Hiroshi Saruwatari</i>	
D2-0830_L1.8 Three-Dimensional Gradient-Based Tracking of Multiple Sound Sources ..	113
<i>Shaoheng Xu, Wei-Ting Lai, Yile (Angela) Zhang, Jihui (Aimee) Zhang, Amy Bastine, Prasanga Samarasinghe, Thushara Abhayapala</i>	
D2-0830_L1.9 Retrieval-Augmented Difference Captioning to Explain Unsupervised Anomalous Sound Detection	119
<i>Ryoya Ogura, Tomoya Nishida, Yohei Kawaguchi</i>	
D2-0830_L1.10 An Evaluation of Supervised Virtual Microphone Estimators in Reverberant Sound Fields	125
<i>Kimihiro Hattori, Wen-Chin Huang, Kazuya Takeda, Tomoki Toda</i>	
D2-0830_L1.11 Human-CLAP: Human-perception-based contrastive language–audio pretraining	131
<i>Taisei Takano, Yuki Okamoto, Yusuke Kanamori, Yuki Saito, Ryotaro Nagase, Hiroshi Saruwatari</i>	
D2-0830_L1.12 Training Acoustic Scene Classification Models Robust to Asynchrony in Distributed Microphone Arrays	137
<i>Takao Kawamura, Nobutaka Ono</i>	
Grand Challenge & Audio Processing	
D3-1030_L1.1 DG-SED: Domain Generalization for Sound Event Detection with Heterogeneous Training Data	143
<i>Yang Xiao, Han Yin, Jisheng Bai, Rohan Kumar Das</i>	
D3-1030_L1.2 Accelerated Convolutional Transfer Function-Based Multichannel NMF Using Iterative Source Steering	149
<i>Xuemai Xie, Xianrui Wang, Liyuan Zhang, Yichen Yang, Shoji Makino</i>	
D3-1030_L1.3 DySiME: Dynamic Single-Source Multichannel Enhancement Using Time-Varying Directional Cues	154
<i>Hao Liang, Yichen Yang, Xiao Zhang, Shoji Makino, Jingdong Chen</i>	

D3-1030_L1.4 Demixing Filter Estimation for Bleeding-Sound Reduction of a Vocal Microphone	160
<i>Soushi Taninomiya, Daichi Kitamura, Norihiro Takamune, Kouei Yamaoka, Hiroshi Saruwatari, Yu Takahashi, Kazunobu Kondo, Hayato Yamakawa</i>	
D3-1030_L1.5 Prior-Guided Source Separation with Direct Update of Back-projected Demixing Vectors	166
<i>Kukuru Koiso, Taishi Nakashima, Nobutaka Ono</i>	
D3-1030_L1.6 Meta-Learning with Pretrained Audio Representations Enables One-Shot Acoustic Signal Classification	172
<i>Haoxiang Wu, Zhengqiao Zhao, Jingdong Chen, Jacob Benesty</i>	
D3-1030_L1.7 A Semi-Supervised Acoustic Scene Classification Network Based on Multi-Modal Information Fusion	177
<i>Junkang Yang, Hongqing Liu, Liming Shi, Lu Gan, Hiromitsu Nishizaki, Chee Siang Leow</i>	
D3-1030_L1.8 ASCMamba: Multimodal Time-Frequency Mamba for Acoustic Scene Classification	182
<i>Bochao Sun, Dong Wang, Zhanlong Yang, Jun Yang, Han Yin</i>	
D3-1030_L1.9 Evaluation of Low-Frequency Restriction, Pitch-Shift Augmentation, and Average Pooling for Acoustic Scene Classification under Unseen-City Conditions	188
<i>Takao Kawamura, Masayuki Sera, Nobutaka Ono</i>	
D3-1030_L1.10 The APSIPA ASC 2025 Grand Challenge on City and Time-Aware Semi-supervised Acoustic Scene Classification: Summary and Results	193
<i>Jisheng Bai, Mou Wang, Haohe Liu, Bin Xiang, Yin Liu, Jianfeng Chen, Dongyuan Shi, Mark Plumbley, Susanto Rahardja, Woon-Seng Gan</i>	
Advanced Topics on Music Processing	
D2-1030_L1.1 Drum-to-Vocal Percussion Sound Conversion and Its Evaluation Methodology	198
<i>Rinka Nobukawa, Makito Kitamura, Tomohiko Nakamura, Shinnosuke Takamichi, Hiroshi Saruwatari</i>	
D2-1030_L1.2 How do Deaf and Hard of Hearing people listen to Music Instruments? Subjective Evaluation and Acoustic Features	204
<i>Rumi Hiraga, Yuhki Shiraishi, Keiichi Yasu</i>	
D2-1030_L1.3 Quality Assessment of DNN-Based Algorithms for Music Boundary Detection	210
<i>Aneeka Azmat, Li Su, ChengHsin Hsu</i>	
D2-1030_L1.4 Note-level Nonchord-tone Identification with Graph Neural Networks	216
<i>Yui Uehara, Satoshi Tojo</i>	

D2-1030_L1.5 Evaluation Score Prediction for Japanese Songs Based on Melody Fitness to Lyrics 222
Sosuke Nishimura, Eita Nakamura

D2-1030_L1.6 A Comparative Study of Statistical Features and Deep Learning for Orchestral Texture Classification 228
Zih-Syuan Lin, Jun-You Wang, Li Su

D2-1030_L1.7 Efficient Transformer-Based Piano Transcription With Sparse Attention Mechanisms 234
Weixing Wei, Kazuyoshi Yoshii

D2-1030_L1.8 Transformer-Based Unpaired Piano Accompaniment Style Transfer 240
Hsin Ai, Yi-Hsuan Yang

D2-1030_L1.9 Designing a Music Difficulty Measure for Controllable Automatic Piano Rearrangement 246
Hikari Miyaji, Keito Sawada, Wen-Chin Huang, Tomoki Toda

D2-1030_L1.10 Vocal onset detection and pitch segmentation in medieval choral music guided by original notational sources 252
Samuel Bellows, Sarabeth Mullins, Brian Katz

D2-1030_L1.11 MORTM: MoE-Optimized Rhythmic Transformer Model for Symbolic MIDI Generation 258
Takaaki Nagoshi, Tetsuro Kitahara

D2-1030_L1.12 TAPA-ICL: Taxonomy-Aware Prompt Augmentation for In-Context Learning in Music Understanding 264
Jiahao Zhao, Yunjia Li, Kazuyoshi Yoshii

Music and Singing

D3-0830_L1.1 Unified Timbre Transfer: A Compact Model for Real-Time Multi-Instrument Sound Morphing 270
Anders R. Bargum, Naotake Masuda, Bogdan Teleaga, Andrew Fyfe, Cumhur Erkut

D3-0830_L1.2 REAL-WORLD MUSIC PLAGIARISM DETECTION WITH MUSIC SEGMENT TRANSCRIPTION SYSTEM 276
Seonghyeon Go

D3-0830_L1.3 Attention-based Adaptive Structured Patchout Spectrogram Transformer for Music Classification 282
Yuan Liu, Shoji Makino, Lingqing Liu, Yichen Yang

D3-0830_L1.4 Accuracy Improvement of Automatic Chord Recognition with Source Separation Preprocessing	288
<i>Ayumu Mitoma, Ken'ichi Furuya</i>	
D3-0830_L1.5 Effects of Music Training Experience on the Production of English Rhythm by Chinese Learners	293
<i>Ying Chen, Chenyu Li, Ruizhe Wang, Yujia Zhang</i>	
D3-0830_L1.6 Hierarchical Symbolic Music Generation with Variational Autoencoder-based Bar-wise Feature Sequences	299
<i>Keito Sawada, Wen-Chin Huang, Tomoki Toda</i>	
D3-0830_L1.7 Singing MIDI Transcription with Music Language Models: Formulation and Comparison	305
<i>Yu Sugimoto, Jun-You Wang, Li Su, Eita Nakamura</i>	
D3-0830_L1.8 Data-Efficient Music Captioning via Contrastive and Semantic Alignment	311
<i>Leekyung Kim, Jonghun Park</i>	
D3-0830_L1.9 GAN-Enhanced InpaintNet for Music Inpainting on Limited Data	317
<i>Koumei Naemura, Boyu Cao, Ryotaro Nagase, Ryoichi Takashima, Yoichi Yamashita</i>	
D3-0830_L1.10 An Analysis of Singing Accuracy towards Quantifying the Melodic Singability	323
<i>Minami Kawahara, Tetsuro Kitahara</i>	
D3-0830_L1.11 Guitar Tone Morphing by Diffusion-based Model	329
<i>Kuan-Yu Chen, Kuan-Lin Chen, Yu-Chieh Yu, Jian-Jiun Ding</i>	
Active Noise Control I	
D2-1430_L1.1 Design of speech leakage-suppressed audio-spot based on auditory masking area control with active masker cancellation using parametric array loudspeakers	334
<i>Tomoki Hashida, Yuting Geng, Masato Nakayama, Takano bu Nishiura</i>	
D2-1430_L1.2 Multichannel feedforward active noise control system with optical laser microphone in reverberant environments	340
<i>Maoto Mizutani, Kenta Iwai, Masato Nakayama, Takano bu Nishiura, Yoshiharu Soeta</i>	
D2-1430_L1.3 Frequency-domain online modeling of multiple secondary paths without auxiliary noise for active noise control	346
<i>Siyuan Lian, Xiaofeng Zeng, Ruquan Sun, Jing Lu</i>	
D2-1430_L1.4 Applying Model-Agnostic Meta-Learning with Iterative Dichotomiser 3 for Alternating-Switching Active Noise Control Systems	352
<i>Xiaoyi Shen, Dongyuan Shi, Woon-Seng Gan, Jun Yang</i>	

D2-1430_L1.5 A Robust Proactive Communication Strategy for Distributed Active Noise Control Systems	358
<i>Junwei Ji, Dongyuan Shi, Zhengding Luo, Boxiang Wang, Ziyi Yang, Haowen Li, Woon-Seng Gan</i>	
D2-1430_L1.6 Directional Selective Fixed-Filter Active Noise Control Based on a Convolutional Neural Network in Reverberant Environments	364
<i>Boxiang Wang, Zhengding Luo, Haowen Li, Dongyuan Shi, Junwei Ji, Ziyi Yang, Woon-Seng Gan</i>	
D2-1430_L1.7 An Online Secondary Path Modeling Technique in a Hybrid Active Noise Control System	370
<i>Harold Alexis Lao, Cheng-Yuan Chang</i>	
D2-1430_L1.8 A Diffusion Remote Microphone Technique for Distributed Active Noise Control	376
<i>Tianyou Li, Sipei Zhao, Haowen Li, Xiaofeng Zeng, Ruquan Sun, Jing Lu</i>	
D2-1430_L1.9 An Integrated Active Noise Control and Crosstalk Cancellation System Designed Under a Generalized Model-Matching Framework	382
<i>Michael Edy, Chih Yen Wang, Ching En Huang, You Siang Chen, Mingsian R. Bai</i>	
D2-1430_L1.10 Improvement of Noise Reduction in a Panel Combined with Multiple Loudspeakers Using Active Noise Control	388
<i>Tatsuya Murao</i>	
D2-1430_L1.11 Selective Fixed Filter Sub-band Active Noise Control System Based on Reference Signal Power Estimation	394
<i>Shota Toyooka, Ryo Matsuura, Kenta Iwai, Yoshinobu Kajikawa</i>	
D2-1430_L1.12 Performance analysis of active noise control over a spatial region	400
<i>Jihui (Aimee) Zhang, Thushara Abhayapala, Naoki Murata, Prasanga Samarasinghe, Yu Maeno, Yuki Mitsufuji</i>	
Active Noise Control II	
D2-1630_L1.1 Electro-acoustic component placement optimization for helicopter cabin ANC systems	406
<i>Yuhang Yang, Liquan Shi, Ningyuan Liang, Guoyong Jin</i>	
D2-1630_L1.2 Spatial-Correlation-Based Error Weighting Method for Efficient Application of Filtered Reference Algorithm in Multichannel Active Noise Control	412
<i>Meiling Hu, Jing Lu, Qingyu Ma</i>	
D2-1630_L1.3 An Alternating Mode Strategy for Adaptive Sound Field Control and Acoustic Path Tracking	417
<i>Junqing Zhang, Jingli Xie, Dongyuan Shi, Wen Zhang, Jingdong Chen, Jacob Benesty</i>	

D2-1630_L1.4 DOA Estimation with Lightweight Network on LLM-Aided Simulated Acoustic Scenes 423
Haowen Li, Zhengding Luo, Dongyuan Shi, Boxiang Wang, Junwei Ji, Ziyi Yang, Woon-Seng Gan

D2-1630_L1.5 Unsupervised Spectrogram Enhancement Algorithm Based on Bi-LSTM ... 429
Hanwen Zhang, Xiruo Su, Zhijuan Zhu, Bin Wu, Lingyun Ye

D2-1630_L1.6 Continual Learning-Based Selective Fixed-filter Active Noise Control 435
Jingsong Xiao, Qirui Huang

D2-1630_L1.7 Meta-Learned Regional Initialization of Control Filters for Headphone Active Noise Control 441
Ziyi Yang, Zhengding Luo, Dongyuan Shi, Junwei Ji, Boxiang Wang, Haowen Li, Qirui Huang, Woon-Seng Gan

D2-1630_L1.8 RAMDC: Room-Aware Multi-Device Clustering for Large Scale Teleconferencing 447
Yile Zhang, Weiting Lai, Amy Bastine, Xingyu Chen, Lachlan Birnie, Thushara Abhayapala, Prasanga Samarasinghe

D2-1630_L1.9 Multi-channel ANC with Adaptive Kernel Assisted On-line Secondary Path Modeling 453
Hucheng Wang, Tao Liu, Junqing Zhang, Wen Zhang

D2-1630_L1.10 A Laplace Distribution-Based Variable Step-Size FxlogLMS Algorithm for Active Impulsive Noise Control 459
Aoi Haneda, Yosuke Sugiura, Tetsuya Shimamura

D2-1630_L1.11 Research Progress on Active Control of Road Noise in Vehicles 465
Wangxiaoxu Chen, Jiancheng Tao, Shuping Wang, Kai Chen, Haishan Zou, Xiaojun Qiu

Audio Processing

D2-1430_P1.1 Anomalous Sound Detection Based on Derivative Features of Short-Time Holomorphic Fourier Transform 471
Iori Hashimoto, Yu Morinaga, Suehiro Shimauchi, Shigeaki Aoki

D2-1430_P1.2 Elastic Additive Angular Margin Loss Integrated with Mixup for Anomalous Sound Detection 477
Yihao Zhao, Yichen Yang, Xiao Zhang, Shoji Makino

D2-1430_P1.3 A Distilled Low-Latency Neural Vocoder with Explicit Amplitude and Phase Prediction 483
Hui-Peng Du, Yang Ai, Zhen-Hua Ling

D2-1430_P1.4 Directional Filtering of Sound Fields for Emphasizing Specific Directions of Arrival and Its Applications	489
<i>Ryo Murakami, Natsuki Ueno</i>	
D2-1430_P1.5 Sound Field Estimation Method Robust to Microphone Position and Directivity Errors	495
<i>Takumi Koga, Natsuki Ueno</i>	
D2-1430_P1.6 Anomalous Sound Detection Using Time-Frequency Derivative of Instantaneous Phase Features	501
<i>Tran-Quang-Tuan Vo, Quoc-Huy Nguyen, Masashi Unoki</i>	
D2-1430_P1.7 Few-Step Diffusion-Based Voice Conversion Using Consistency Trajectory Models	507
<i>Ryuichi Hatakeyama, Toru Nakashika, Takuya Takahashi</i>	
D2-1430_P1.8 Spatial Audio Signal Enhancement: A Multi-output MVDR Method in The Spherical Harmonic-domain	513
<i>Huawei Zhang, Jihui (Aimee) Zhang, Huiyuan (June) Sun, Prasanga Samarasinghe</i>	
Interactive, Natural, Expressive and Robust Conversation System	
D1-1430_L2.1 Language Adaptation Wake Word Spotting via Latent Space from Pre-trained Speech Models	519
<i>Shifu Xiong, Hengshun Zhou, Kai Shen, Shi Cheng, Hang Chen, Genshun Wan, Kewei Li, Jun Du, Lirong Dai</i>	
D1-1430_L2.2 Identifying Speaker Information in Feed-Forward Layers of Self-Supervised Speech Transformers	525
<i>Tzu-Quan Lin, Hsi-Chun Cheng, Hung-yi Lee, Hao Tang</i>	
D1-1430_L2.3 Multi-task Pretraining for Enhancing Interpretable L2 Pronunciation Assessment	531
<i>Jiun-Ting Li, Bi-Cheng Yan, Yi-Cheng Wang, Berlin Chen</i>	
D1-1430_L2.4 End-to-End Integration of Speech Emotion Recognition and Voice Activity Detection with a Self-Supervised Model for Noise Robustness	537
<i>Natsuo Yamashita, Masaaki Yamamoto, Yohei Kawaguchi</i>	
D1-1430_L2.5 SCSMT: A Multilingual Children's Speech Corpus for Singapore's Mother Tongues	543
<i>Bowen Zhang, Nur Afiqah Abdul Latiff, Rong Tong, Donny Soh, Ian McLoughlin</i>	
D1-1430_L2.6 Reducing Orthographic Dependency on Paired Data by Probabilistic Integration via Syllabogram for Japanese Dialogue Speech Recognition	549
<i>Ryu Takeda, Kazunori Komatani</i>	

D1-1430_L2.7 Expressive Prompting: Improving Emotion Intensity and Speaker Consistency in Zero-Shot TTS	555
<i>Haoyu Wang, Chunyu Qiang, Tianrui Wang, Cheng Gong, Yu Jiang, Yuheng Lu, Chen Zhang, Longbiao Wang, Jianwu Dang</i>	
D1-1430_L2.8 Constructing an In-the-Wild Spoken Dialogue Dataset Based on YouTube Dialogue Videos	561
<i>Yuki Sato, Sanae Yamashita, Shinnosuke Takamichi, Ryuichiro Higashinaka</i>	
D1-1430_L2.9 Emotional Text-To-Speech Based on Mutual-Information-Guided Emotion-Timbre Disentanglement	567
<i>Jianing Yang, Sheng Li, Takahiro Shinozaki, Yuki Saito, Hiroshi Saruwatari</i>	
D1-1430_L2.10 Conversation Context-aware Direct Preference Optimization for Style-Controlled Speech Synthesis	573
<i>Atsushi Kojima, Yusuke Fujita, Hao Shi, Tomoya Mizumoto, Mengjie Zhao, Yui Sudo</i>	
D1-1430_L2.11 A Hybrid Attention Mechanism to Improve Tacotron 2 Performance for Indonesian Text-to-Speech Synthesis	579
<i>Angela Catherina, Bima Prihasto, Bobby Mugi Pratama, Li-Wei Kang, Jia-Ching Wang</i>	
Speaker Modeling Beyond Speaker Recognition	
D1-1630_L2.1 SpkAugTSE: A Simple and Efficient Approach to Address Target Confusion in End-to-End Speaker Extraction	583
<i>Zhenghai You, Zhenyu Zhou, Lantian Li, Dong Wang</i>	
D1-1630_L2.2 Interpolating Speaker Identities in Embedding Space for Data Expansion .	589
<i>Tianchi Liu, Ruijie Tao, Qiongqiong Wang, Yidi Jiang, Hardik B. Sailor, Ke Zhang, Jingru Lin, Haizhou Li</i>	
D1-1630_L2.3 MDD: a Mask Diffusion Detector to Protect Speaker Verification Systems from Adversarial Perturbations	595
<i>Yibo Bai, Sizhou Chen, Michele Panariello, Xiao-Lei Zhang, Massimiliano Todisco, Nicholas Evans</i>	
D1-1630_L2.4 Fusing Multi-layer Features of the Pre-trained Model With Grouped Cross Attention for Spoofing Speech Detection	601
<i>Yu Guan, Wu Guo, Jie Zhang, Zhijun Zhang</i>	
D1-1630_L2.5 Fusing Blocked Deep Features of Pre-Trained Models for Short-Duration Speaker Verification	607
<i>Zhi jun Zhang, Wu Guo, Jie Zhang, Yu Guan</i>	
D1-1630_L2.6 Multi-level Adversarial Training with Data Augmentation for Robust Speaker Verification	613
<i>Xiaolei Zhang, Zhihua Fang, Liang He</i>	

D1-1630_L2.7 Analysis of Speaker Verification Performance Trade-offs with Neural Audio Codec Transmission 619
Nirmalya Mallick Thakur, Jia Qi Yip, Eng Siong Chng

D1-1630_L2.8 Estimating Speaker’s Seating Position from Monaural Speech in a Simulated Vehicle Interior Sound Field 625
Masataka Kaneko, Wen-Chin Huang, Tomoki Toda

D1-1630_L2.9 TS-VAD+: Modularized Target-Speaker Voice Activity Detection for Robust Speaker Diarization 630
Tran The Anh, Azmat Adnan, Wu Yihao, Chng Eng Siong

D1-1630_L2.10 Are Multimodal Foundation Models All That Is Needed for Emofake Detection? 636
Mohd Mujtaba Akhtar, Girish, Orchid Chetia Phukan, Swarup Ranjan Behera, Pailla Balakrishna Reddy, Ananda Chandra Nayak, Sanjib Kumar Nayak, Arun Balaji Buduru

Speech and Language Processing I

D2-0830_L2.1 Neural Speech Separation with Parallel Amplitude and Phase Spectrum Estimation 642
Fei Liu, Yang Ai, Zhen-Hua Ling

D2-0830_L2.2 Single-Channel Speech Enhancement in Spherical-Mapped Short-Time Spectral Domain 648
Yu Morinaga, Naoto Kotake, Iori Hashimoto, Suehiro Shimauchi, Shigeaki Aoki

D2-0830_L2.3 Introducing Self-Supervised Learning Models for Spoken Query-Spoken Term Detection 653
Masato Nagase, Kazunori Kojima, Shi-wook Lee, Yoshiaki Itoh

D2-0830_L2.4 Characterization of Speech Similarity Between Australian Aboriginal and High-Resource Languages: A Case Study on Dharawal 658
Ting Dang, Trini Manoj Jeyaseelan, Eliathamby Ambikairajah, Vidhyasaharan Sethu

D2-0830_L2.5 Segment Transformer: AI-Generated Music Detection via Music Structural Analysis 664
Yumin Kim, Seonghyeon Go

D2-0830_L2.6 Dialect Identification Using Resource-Efficient Fine-Tuning Approaches ... 670
Zirui Lin, Haris Gulzar, Monnika Roslianna Busto, Akiko Masaki, Takeharu Eda, Kazuhiro Nakadai

D2-0830_L2.7 A High-Quality and Low-Complexity Streamable Neural Speech Codec with Knowledge Distillation 676
En-Wei Zhang, Hui-Peng Du, Xiao-Hang Jiang, Yang Ai, Zhen-Hua Ling

D2-0830_L2.8 Effectiveness of streaming ASR for real-time laughter and screaming detection 682
Mizuki Kurasawa, Yoshiko Arimoto

D2-0830_L2.9 Mitigating Data Imbalance in Automated Speaking Assessment 688
Fong-Chun Tsai, Kuan-Tang Huang, Bi-Cheng Yan, Tien-Hong Lo, Berlin Chen

D2-0830_L2.10 An Information-Theoretic Approach to Data Selection for Generative Topic Modeling 694
Michael Santoso, Bhone Tay Zar Kyaw, Valentinus Roby Hananto, Victor Kryssanov

D2-0830_L2.11 Collective Learning-based Optimal Transport GAN with Multi-Level Fine-Grained and Global Discriminators for Voice Conversion 700
Sandipan Dhar, Md. Tousin Akhter, Nanda Dulal Jana, Swagatam Das, Monorama Swain, Saurav Chowdhury

Speech and Language Processing II

D2-1030_L2.1 Beyond Binary Detection: Multi-Etiology Dysarthria Classification with Pre-trained Speech Models 706
Zihan Zhong, Qianli Wang, Satwinder Singh, Clarion Mendes, Mark Hasegawa-Johnson, Waleed Abdulla, Seyed Reza Shahamiri

D2-1030_L2.2 A Dual-Path Speaker-Independent Acoustic-to-Articulatory Inversion Model Based On Content and Speaker Information Disentanglement 712
Qiang Fang

D2-1030_L2.3 Dementia Prediction From Speech Signal Using Optimized Prosodic Features 718
Bagus Tris Atmaja, Sakriani Sakti

D2-1030_L2.4 Speech Emotion Recognition via Entropy-Aware Score Selection 724
ChenYi Chua, JunKai Wong, Chengxin Chen, Xiaoxiao Miao

D2-1030_L2.5 Improving Exemplar-Based Electrolaryngeal Speech Voice Conversion via Robust Content Representations 730
Fo-Rui Li, Hsin-Te Hwang, Ming-Chi Yen, Men-Tung Lo, Yu Tsao, Hsin-Min Wang

D2-1030_L2.6 An Efficient Transfer Learning Method Based on Adapter with Local Attributes for Speech Emotion Recognition 736
Haoyu Song, Mcloughlin Ian, Qing Gu, Nan Jiang, Yan Song

D2-1030_L2.7 ASRQ-VC: ASR-Guided Speech Content Quantization for High-Fidelity Voice Conversion 741
Songting Liu, Deheng Ye, Wei Yang, Haoyang Li, Eng Siong Chng

D2-1030_L2.8 PUNSER: Large-Scale Pre-trained and Unified Model for Practical Speech Emotion Recognition 747
Yu Hayashizaki, Takashi Nose, Sumiharu Kobayashi, Satoru Fukayama, Akinori Ito

D2-1030_L2.9 Investigation of the effectiveness of converted speech auditory feedback in low-latency real-time voice conversion 753
Kiseki Niwa, Kazuhiro Kobayashi, Tomoki Toda

D2-1030_L2.10 Study on Signal Processing Techniques in Protecting Voice Personae Against Speech Synthesis Systems 759
Nopparut Li, Candy Olivia Mawalim, Masashi Unoki

D2-1030_L2.11 MixedG2P-T5: G2P-free Speech Synthesis for Mixed-script texts using Speech Self-Supervised Learning and Language Model 765
Joonyong Park, Daisuke Saito, Nobuaki Minematsu

Speech and Language Processing III

D2-1430_L2.1 I²TTS: Image-indicated Immersive Text-to-speech Synthesis with Spatial Perception 771
Jiawei Zhang, Tian-Hao Zhang, Jun Wang, Jiaran Gao, Ruijie Tao, Xinyuan Qian, Xu-Cheng Yin

D2-1430_L2.2 Chain-of-Thought Distillation for ASR Error Correction with Multimodal Large Language Models 777
Shaomeng Yang, Jiaming Luo, Jinran Wang, Rongfeng Su, Yongjie Zhou, Lan Wang, Nan Yan

D2-1430_L2.3 Direction-guided Spatial Attention for Multichannel Speech Enhancement 783
Shuai Nie, Yaran Chen, Shan Liang, Jiaming Xu, Runyu Shi

D2-1430_L2.4 A Study of Japanese Mixed Emotional Speech Synthesis Based on an End-to-End Emotional Speech Synthesis Model 789
Issei Sakata, Tetsuo Kosaka

D2-1430_L2.5 EFTTS: Zero-Shot Emotional Speech Synthesis via Conditional Flow Matching and Self-Supervised Representations 795
Haoyu Wang, Jiale Chen, Jiaxun Li, Sizhe Shan, Yuehai Wang

D2-1430_L2.6 Improving Speech-to-Speech Translation for Low-Resource Languages via Transfer Learning 801
Rui Zhou, Akinori Ito, Takashi Nose

D2-1430_L2.7 DialoSpeech: Dual-Speaker Dialogue Generation with LLM and Flow Matching 807
Hanke Xie, Dake Guo, Chengyou Wang, Yue Li, Wenjie Tian, Xinfu Zhu, Xinsheng Wang, Xiulin Li, Guanqiong Miao, Bo Liu, Lei Xie

D2-1430_L2.8 VICNet: FaderNet-Based Voice Impression Conversion with Affective Dimensional Representation	813
<i>Takuya Takahashi, Saki Kugimoto, Toru Nakashika</i>	
D2-1430_L2.9 Strategic Re-weighting of U-Net Components in Diffusion Models for Enhanced Speech Enhancement without Retraining	819
<i>Yuehai Zhang, Yang Li, Yuehao Zhao, Shoji Makino</i>	
D2-1430_L2.10 Fast and Speaker-Independent Utterance Selection for ASR-Free CALL Systems of Minority Languages	825
<i>Takaki Koshikawa, Akinori Ito, Takashi Nose</i>	
D2-1430_L2.11 Speech-Content-Driven Highlighting of Translated Lecture Slides for Foreign Language Lecture Understanding	831
<i>Naoki Muto, Chee Siang Leow, Junichi Hoshino, Takehito Utsuro, Hiromitsu Nishizaki</i>	
Speech and Language Processing IV	
D2-1630_L2.1 Leveraging Language Information for Target Language Extraction	837
<i>Mehmet Sinan Yildirim, Ruijie Tao, Wupeng Wang, Junyi Ao, Haizhou Li</i>	
D2-1630_L2.2 VietLyrics: A Large-Scale Dataset and Models for Vietnamese Automatic Lyrics Transcription	843
<i>Nguyen Quoc Anh, Bernard Cheng, Kelvin Soh</i>	
D2-1630_L2.3 Autofocus Neural Beamformer Based on Steering Vector Estimation	849
<i>Reiya Marukawa, Takeshi Yamada</i>	
D2-1630_L2.4 Estimating User Sentiment at Sub-exchange Granularity from Exchange-level Annotations	855
<i>Daichi Yukizawa, Kazunori Komatani, Ryu Takeda, Kenta Yamamoto</i>	
D2-1630_L2.5 DAU-KDAH Dysarthic Multi-Lingual and Multimodal Speech Corpora for Indic Languages	861
<i>Arth Shah, Hiya Chaudhari, Kavya Kumar, Arushi Srivastava, Priya Damdar, Ravindrakumar Purohit, Dharmendra Vaghera, Bhavna Singh, Aparna Walanj, Abhishek Srivastava, Hemant Patil</i>	
D2-1630_L2.6 Gamma-VAE-VC: Voice conversion based on VAE assuming gamma distribution for both latent variables and observation	867
<i>Nanako Imaichi, Takuya Takahashi, Toru Nakashika</i>	
D2-1630_L2.7 Zero-shot Context Biasing with Trie-based Decoding using Synthetic Multi-Pronunciation	873
<i>Changsong Liu, Yizhou Peng, Eng Siong Chng</i>	

D2-1630_L2.8 Dimension 414 and Minimal Embedding Dimensions for Phonetic Feature Encoding in WavLM 879
Narthana Sivalingam, Uthayasanker Thayasivam

D2-1630_L2.9 Directional Hybrid Optimization of HRTFs for Low-Order Spherical Harmonics Binaural Rendering 885
Rui Zhang, Yuxuan Ke, Qunping Ni, Ge Yao, Xiaodong Li, Chengshi Zheng

D2-1630_L2.10 Speech Enhancement Network With Windowed Cross Attention Using Noise-Reference Microphone 891
Kota Suzuki, Yosuke Sugiura, Tetsuya Shimamura

D2-1630_L2.11 BAANI: A 296M-Parameter Neural Vocoder for End-to-End Punjabi Speech Synthesis 897
Siddharth Kumar, Nisarg Trivedi, Ravindrakumar Purohit, Hemant Patil

Speech Communication and Media Generation

D3-0830_L2.1 Active Learning for Text-to-Speech Synthesis with Informative Sample Collection 903
Kentaro Seki, Shinnosuke Takamichi, Takaaki Saeki, Hiroshi Saruwatari

D3-0830_L2.2 Semi-Supervised End-to-End Speech-to-Text Translation with Joint Text-to-Text and Speech-to-Text Decoding 909
Tomohiro Tanaka, Ryo Masumura, Naoki Makishima, Mana Ihori, Shota Orihashi, Satoshi Suzuki, Taiga Yamane

D3-0830_L2.3 UTRo-NAST: Non-Autoregressive Speech Translation via Understanding, Translation, and Reordering 915
Yu-Chen Kuan, Kuan-Yu Chen

D3-0830_L2.4 Laughing Across Borders: A Culturally-Aware Joke Generator for Asian Regions 921
Ashley Fang Cai Xian, Chen Ting Ng, Ashley Kok Siu Cheng, Wah Yang Tan, Mohan Raj Chanthran, Lay-Ki Soon, Meisin Lee

D3-0830_L2.5 Synthesizing Vowel-Like Tones with Pitch Circularity 926
Kaori Hashimoto, Takao Kawamura, Nobutaka Ono

D3-0830_L2.6 Error Correction Using LLMs for Sentence Estimation from Ambiguous Inputs via Wearable Keyboards 932
Matsuri Iwasaki, Masanobu Abe, Sunao Hara

D3-0830_L2.7 A Robust End to End Spoken Grammar Assessment System 938
Sunil Kumar Kopparapu, Chitrlekha Bhat, Ashish Panda

D3-0830_L2.8 LAPS-Diff: A Diffusion-Based Framework for Hindi Singing Voice Synthesis With Language Aware Prosody-Style Guided Learning 944
Sandipan Dhar, Mayank Gupta, Preeti Rao

D3-0830_L2.9 End-to-end multi-channel speaker extraction and binaural speech synthesis 950
Cheng Chi, Xiaoyu Li, Yuxuan Ke, Qunping Ni, Ge Yao, Xiaodong Li, Chengshi Zheng

D3-0830_L2.10 Improving Listening Head Generation Performance Using Speech Representations from Self-Supervised Learning 956
Tamon Mikawa, Yasuhisa Fujii, Yukoh Wakabayashi, Kengo Ohta, Ryota Nishimura, Norihide Kitaoka

D3-0830_L2.11 ULF-TTS: An Uncluttered Hybrid TTS System using Language and Flow Matching Models 962
Jae Hyun Park, Young Sik Eom, SeungJae Choi, Allison Shindell, Min-Gwan Seo, Gyeong-hoon Lee

Speech Recognition and Understanding I

D3-1030_L2.1 Phoneme-grapheme Dictionary-based Prompting for Robust Proper Noun Recognition in Japanese ASR 968
Ryuga Sugano, Hiroaki Sato, Asahi Sakuma, Tadashi Kumano, Yoshihiko Kawai, Shinji Watanabe

D3-1030_L2.2 LLM-Driven Hypothesis Set Refinement for Enhanced ASR Post-Processing 974
Chen-Han Wu, Kuan-Yu Chen

D3-1030_L2.3 Real-time VAD-less speech recognition by fine-tuning SSL model with data containing tagged non-speech segments 980
Jotaro Emoto, Ryota Nishimura, Kengo Ohta, Norihide Kitaoka

D3-1030_L2.4 Improving Automatic Speech Recognition Model for Super-Elderly Voice Using Speech Synthesis Model 986
Ryota Uematsu, Chee Siang Leow, Norihide Kitaoka, Hiromitsu Nishizaki

D3-1030_L2.5 Improving Code-Switching Speech Recognition with TTS Data Augmentation 992
Yue Heng Yeo, Yu Chen Hu, Shreyas Gopal, Yi Zhou Peng, He Xin Liu, Eng Siong Chng

D3-1030_L2.6 PQSR: A Speech Corpus of Polar Questions and Spontaneous Responses in Standard Chinese with Complex Intentions Annotated 998
Yingyi Luo, Yue Huang, Qingke Sun, Shuwen Chen

D3-1030_L2.7 Toward Natural System Repair: An Analysis of Human Other-Initiated Self-Repair Patterns in Japanese Casual Conversations 1004
Kazuya Tsubokura, Yurie Iribe, Norihide Kitaoka

D3-1030_L2.8 Self-Supervised Learning for Classification of Normal vs. Dysarthric Speech	1010
<i>Hiya Chaudhari, Kavya Kumar, Hemant Patil</i>	
D3-1030_L2.9 Investigating Polyglot Speech Foundation Models for Learning Collective Emotion from Crowds	1016
<i>Orchid Chetia Phukan, Girish, Mohd Mujtaba Akhtar, Panchal Nayak, Priyabrata Mallick, Swarup Ranjan Behera, Parabattina Bhagath, Pailla Balakrishna Reddy, Arun Balaji Buduru</i>	
D3-1030_L2.10 Rethinking Cross-Corpus Speech Emotion Recognition Benchmarking: Are Paralinguistic Pre-Trained Representations Sufficient?	1022
<i>Orchid Chetia Phukan, Mohd Mujtaba Akhtar, Girish, Swarup Ranjan Behera, Parabattina Bhagath, Pailla Balakrishna Reddy, Arun Balaji Buduru</i>	
Speech Recognition and Understanding II	
D3-1430_L2.1 Probabilistic Language-Aware Speech Recognition	1028
<i>Jen-Tzung Chien, Willianto Sulaiman, Chung-Hsuan Wang</i>	
D3-1430_L2.2 LLMs-Integrated Automatic Hate Speech Recognition Using Controllable Text Generation Models	1033
<i>Ryutaro Oshima, Yuya Hosoda, Yoji Iiguni</i>	
D3-1430_L2.3 Multi-stage Speech Enhancement with Cascaded SNR Domain Shifts	1039
<i>Xiaoran Li, Zilu Guo, Jun Du</i>	
D3-1430_L2.4 Autoencoder-Driven Latent Representation Learning for Language-Agnostic Disordered Speech Classification using a Universal Feature set	1045
<i>Puneet Bawa, Virender Kadyan, Shareef Babu Kalluri</i>	
D3-1430_L2.5 FH-RestoreASR: Frequency-Hopping Robust Air Traffic Control Speech Restoration and Recognition	1051
<i>Youngeun Kwon, Yeri Byun, Hyunsung Cho, Jongwon Choi</i>	
D3-1430_L2.6 Speech Intelligibility Assessment with Uncertainty-Aware Whisper Embeddings and sLSTM	1057
<i>Ryandhimas Zezario, Dyah Wisnu, Hsin-Min Wang, Yu Tsao</i>	
D3-1430_L2.7 GCI detection and glottal wave estimation based on TV-CAR speech analysis	1062
<i>Keiichi Funaki</i>	
D3-1430_L2.8 HIPA-MoE: A Parameter-Efficient Fine-Tuning Architecture with Hierarchical Adapter-based Mixture-of-Experts for Multilingual ASR	1068
<i>Xun Lu, Xuyang Wang, Gaofeng Cheng, Lin Zheng, Pengyuan Zhang</i>	

D3-1430_L2.9 Mild Cognitive Impairment Detection via Linear Discriminant Analysis of Picture Description Speech Features: A Cross Corpus Comparison 1074
Yan-Lin Lai, Erh-Yun Chang, Yi-Wen Liu, Jung Lung Hsu, Hui-Chuan Hsu

D3-1430_L2.10 Parameter-Efficient Fine-Tuning of Foundation Models for CLP Speech Classification 1080
Susmita Bhattacharjee, Jagabandhu Mishra, H.S. Shekhawat, S R Mahadeva Prasanna

D3-1430_L2.11 Language Awareness in Code-Switching Speech Recognition 1086
Jen-Tzung Chien, Bobbi Aditya

Speech Synthesis, Enhancement, and Recognition

D3-1430_L1.1 End-to-End Simultaneous Dysarthric Speech Reconstruction with Frame-Level Adaptor and Multiple Wait-k Knowledge Distillation 1092
Minghui Wu, Haitao Tang, Jiahuan Fan, Ruizhi Liao, Yanyong Zhang

D3-1430_L1.2 Disfluency Disentanglement Enhancement in Spoken-Text-Style Transfer for Spontaneous Speech Synthesis 1098
Yuuto Nakata, Daiki Yoshioka, Wen-Chin Huang, Tomoki Toda

D3-1430_L1.3 DARS: Dysarthria-Aware Rhythm-Style Synthesis for ASR Enhancement . 1104
Minghui Wu, Xuelling Liu, Jiahuan Fan, Haitao Tang, Yanyong Zhang, Yue Zhang

D3-1430_L1.4 Emotion-Rich Cross-Speaker TTS via Contrastive Prosody Enhancement 1110
Jen-Tzung Chien, Bryan Gautama Ngo

D3-1430_L1.5 Face-conditioned Large-scale Text-to-Speech via Speaker Embedding Prediction from Facial Images 1116
Umi Okamoto, Sei Ueno, Akinobu Lee

D3-1430_L1.6 Few-shot Speaker Adaptation for Text-to-Speech Synthesis Using Non-Target Speaker Corpora for Glossectomy Patients 1122
Masayori Okamura, Masanobu Abe, Sunao Hara

D3-1430_L1.7 Personalized Bone-Conduction Bandwidth Extension with Speaker Characteristics 1128
Pan Xu, Zhongyu Zhang, Zhonghua Fu

D3-1430_L1.8 Time-Aligned Laughter Sound Event Recognition for Conversational Laughter Analysis and Synthesis 1134
Hiroki Mori

D3-1430_L1.9 PALGAN: A Joint Optimization-Based Preprocessing Method for Speech Restoration in Parametric Array Loudspeakers 1140
Wenyao Ma, Jun Yang

D3-1430_L1.10 Beyond One-Shot Dubbing: Leveraging N-Best Translation and Prompted Paraphrasing with Synchrony-Aware Re-Ranking 1146
Jan Saragih, Faisal Mehmood, Sakriani Sakti

Deep Learning: Algorithm, Implementations, and Applications

D3-0830_H3.1 Honey Adulteration Detection via Robust Diffusion Classifier and Hyperspectral Imaging 1152
Weihao Tang, Guyang Zhang, Waleed Abdulla

D3-0830_H3.2 Semantic-Fast-SAM: Efficient Semantic Segmenter 1158
Byunghyun Kim

D3-0830_H3.3 Micro-expression Recognition Using VideoMamba and Regional Selective Mixup 1164
Yu-Chen Lin, Yi-Jing Chen, Chih-Chang Yu, Hsu-Yung Cheng

D3-0830_H3.4 100x Monolingual Data Augmentation Using LLMs to Build a Parallel Corpus for Machine Translation 1170
Hitoshi Ito, Naoto Shirai, Kazutaka Kinugawa, Hideya Mino, Yoshihiko Kawai

D3-0830_H3.5 Enhancing Technical Documents Retrieval for RAG 1176
Songjiang Lai, Tsun-Hin Cheung, Ka-Chun Fung, Kaiwen Xue, Kwan-Ho Lin, Yan-Ming Choi, Vincent Ng, Kin-Man Lam

D3-0830_H3.6 Lightweight Zero-Shot Keyword Spotting via Multi-Granular Knowledge Distillation 1182
Yun-Ting Sun, Lo-Ya Li, Tien-Hong Lo, Jeh-Weih Hung, Shih-Chieh Huang, Berlin Chen

D3-0830_H3.7 Monomial Matrix Relocation on the Loss Function Level-Set of Feedforward Neural Networks 1188
Ozgur Soysal, Arda Ozdemir, Yiğit Yıldırım, Orhan Arikan

D3-0830_H3.8 Low-Rank Compression of Neural Network Weights by Null-Space Encouragement 1194
Arda Ozdemir, Ozgur Soysal, Ege Doğanay, Yiğit Yıldırım, Orhan Arikan

D3-0830_H3.9 Sign-MExD: An Expert-Infused Diffusion Model for Sign Language Production 1200
Jiayu Shen, Kalin Stefanov, Vee Yee Chong, Lay-Ki Soon, KokSheik Wong

D3-0830_H3.10 FEKF: Flow-based Extended Kalman Filter 1206
Pham Hai Anh, Tran Trong Duy, Do Hai Son, Karim Abed-Meraim, Nguyen Linh Trung

Machine Learning: Algorithms and Application I

D2-1030_P2.1 Kernel Ridge Regression for Efficient Learning of High-Capacity Hopfield Networks 1212
Akira Tamamori

D2-1030_P2.2 Enhanced Sliding Discrete Fourier Transform (eSDFT) with Error-Bound Control for Real-Time Parallel Processing 1218
Jetsada Arnin, Danial Kahani, Bernard A. Conway

D2-1030_P2.3 Sparse-Coded Time-Delay DMD with Control for Nonlinear State-Space Modeling on Graphs 1223
Ryuto Ito, Hiromu Kanauchi, Hiroyasu Yasuda, Masaaki Nagahara, Shogo Muramatsu

D2-1030_P2.4 Nonnegative Matrix Factorization Using Dirichlet-Distribution-Based Regularization 1229
Haru Ogawa, Daichi Kitamura, Shoma Ayano

D2-1030_P2.5 Significance of co-occurring biomarkers in localization of epileptic seizure onset zone 1235
Nawara Mahmood Broti, Masaki Iwasaki, Yumie Ono

D2-1030_P2.6 Reinforcement Learning in Portfolio Management: A Survey of Methods and Trends 1241
Silan Hu, Yulin Huang, Arjun Agarwal, Tanya Warriar, Yuwen Wang, Haozhe Ma, Zhengding Luo

D2-1030_P2.7 Large Sparse Covariance Matrix Estimation via Dual Proximal Gradient Method 1247
Fengpei Li, Ziping Zhao

D2-1030_P2.8 An improved method for Image Shadow Removal by Combining Deterministic and Stochastic Models 1253
Hongjun Sheng, Lanqing Guo, Xinggan Peng, Zhiping Lin, Bihan Wen

D2-1030_P2.9 Knowledge-Infused Topic Model for Empathetic Dialogue Response 1259
Po-Chuan Chen, Jen-Tzung Chien

D2-1030_P2.10 Cross-Patient Seizure Onset Zone Classification by Patient-Dependent Weight 1264
Xuyang ZHAO, Hidenori Sugano, Toshihisa Tanaka

D2-1030_P2.11 NOCTUA: A High-Efficiency Reconfigurable NoC-based Transformer Universal Accelerator 1270
Kun-Chih Chen, Pin-Ching Shen, Bo-Chun Chen

Machine Learning: Algorithms and Application II

D3-0830_P2.1 Equivalence of Graph Signal Processing Using a Hermitian Graph Laplacian and Its Corresponding Graph Laplacian with Duplicated Nodes 1274
Akira Tanaka

D3-0830_P2.2 Skeleton-sequence-based Early Action Recognition by Using Graph Convolutional Neural Networks and Knowledge Distillation Techniques 1279
Wen-Nung Lie, Kien Truc Le, Veasna Vann, Jui-Chiu Chiang, Ngoc Dung Bui

D3-0830_P2.3 A State-Dependent Model for Identification of Time-varying Directed Graphs 1285
Yuzhe Li, Hangjing Zhang, H. Vicky Zhao

D3-0830_P2.4 Unrolled Multimodal Signal Restoration with Signed Twofold Graph Learning 1291
Haruki Yokota, Yuichi Tanaka, Higashi Hiroshi

D3-0830_P2.5 Efficient Sparse Matrix Acceleration for Deep Learning via Two-Step Bitmap Tensor Architecture 1297
Jia-Hong Weng, Tu Wei-Chen

D3-0830_P2.6 Distance-based Laplacian Algebra for Effective Subgraph Filter Learning 1301
Purui Zhang, Feng Ji, Yanan Zhao, Wee Peng Tay, Bihan Wen

D3-0830_P2.7 Evaluation of Low-Resource and High-Efficiency Deep Learning Accelerator for Clinical Dental Diagnosis 1307
Lin Yuan-Jin, Chang Yu-Jen, Liang Chin-Hao, Wei Sung-Tsun, Weng Jia-Hong, Chen Shih-Lun, Tu Wei-Chen

D3-0830_P2.8 Quantization Index Modulation-based Reversible Data Hiding in Compressed Neural Network 1311
Jun Hirano, Jonethe Tan Yang, Fathin Acyuta Makarim, Daham Jayasinghe, KokSheik Wong

D3-0830_P2.9 Dense Vector Retrieval in Data Federation 1317
Amorntip Prayoonwong, Yang-Chun Hsu, Xin-Jie Ye, Po-Kai Lu, Chih-Hang Wang, Chih-Yi Chiu

D3-0830_P2.10 Organ Detection Based on Vision-Language Model For Abdominal CT Images 1323
Jun-Hong Ou, Bo-Xian Wang, Yu-Hong Zheng, Sufal K. Chhabra, Guo-SHiang Lin, Shen-Lei Yan, Chen-Kuo Chiang

D3-0830_P2.11 Algorithm-Architecture Co-Exploration of Systolic Arrays Using High-Level Synthesis 1327
Chu-Chun Yang, Gwo Giun Lee, Tsung-Ying Tsai, Jie-Ren Zheng, Yue-Cong Kuo, Wei-Chieh Lee, Ryan Karthik Pary

Machine Learning: Algorithms and Application III

D3-1030_P2.1 Audio-Visual Fusion Framework for Low-Resource Language Speech Recognition Based on Progressive Down-sampling and Grouped Multi-Heads Attention Mechanism 1332
ChongChong Yu, Xiaolong Xu, Zhaopeng Qian, Kejing Xiao, Yuchen Tan

D3-1030_P2.2 A Data-Driven Control Framework Using Deep Reinforcement Learning for Autonomous Driving 1338
Mei-Lin Huang, Ching-Hung Lee, Cheng-Ting Huang, Hsin-Han Chiang

D3-1030_P2.3 Recipe Diffusion: Cross-Frame Attention and Region-Aware Diffusion for Coherent Visual Recipe Instruction Generation 1344
Weiyi Xia, Satoru Fujita

D3-1030_P2.4 Improving Few-Shot Classification via Feature-Aligned AI-Generated Images 1350
Yu-Wen Tung, Mei-Chen Yeh

D3-1030_P2.5 Rotation Invariant Automatic Rigging for 3D Human Scan Data 1356
Yiqing Li, Satoru Fujita

D3-1030_P2.6 SinDiffPhase: High-Quality Phase Estimation with Ultra-Fast Single-Step Diffusion 1362
Yifei Ni, Andong Li, Lingling Dai, Erwei Yin, Qunping Ni, Chengshi Zheng

D3-1030_P2.7 MapCVAE: Probabilistic Prediction of Diverse Pedestrian Behaviors on General Roads 1368
Konosuke Kobayashi, Satoru Fujita

D3-1030_P2.8 Herald: Democratizing Compositional Reasoning for Visual Tasks without Any Training 1374
Guan Yuan Tan, Arghya Pal, Sailaja Rajanala, Raphael C.-W. Phan, Chee-Ming Ting

D3-1030_P2.9 Canopy to Canopy: Evaluating Model Generalization In 3D Tropical Forest Semantic Segmentation 1380
Sue Han Lee, Brenda Ru Yi Sim, Chung Siung Choo, Yuen Peng Loh

D3-1030_P2.10 LSTM-Transformer Hybrid Network for UAV-Bird Classification Using Radar Track Information 1386
Anning Jiang, Dianfeng Qiao, Shun Liu, Yan Liang

D3-1030_P2.11 A Unified Framework for Interpretable and Uncertainty-Aware Battery State of Health Estimation Using Deep Neural Networks 1392
Elias Isaac Huai-En Lim, Nicholas Heng Loong Wong

Machine Learning: Information and Medical Applications

D2-1430_H3.1 Class Incremental Learning using Continual Backpropagation on Honey Botanic Origin Classification with Hyperspectral Imaging 1398
Guyang Zhang, Iman Ardekani, Waleed Abdulla

D2-1430_H3.2 Multi-strategy improved electric eel foraging optimisation algorithm for UAV path planning 1404
Zexin Zhang, Chengbiao Fu, Hongwei Guo, Anhong Tian

D2-1430_H3.3 A Deep Reinforcement Learning Approach to Roundabout Traffic Signal Control 1412
Cheng-Yu Chen, Daniil Buryakov, Valentinus Roby Hananto, Victor Kryssanov

D2-1430_H3.4 A preliminary study on machine learning to predict circuit exchange in pediatric patients with ECMO 1418
Tatsuya Hasegawa, Toshiyuki Nakanishi, Koichi Fujiwara

D2-1430_H3.5 HasRL Robot: A Heterogeneous Asynchronous Reinforcement Learning System for High-Dimensional Bipedal Control 1423
Jingyang Mai, Zechen Guo, Zhengding Luo, Haozhe Ma

D2-1430_H3.6 A Psychological Strategy Annotation Method Using Multiple LLMs with a Chain of Thought Based on Deductive Reasoning 1429
Jinran Wang, Jiaming Luo, Shaomeng Yang, Yongjie Zhou, Xuefang Zhang, Rongfeng Su, Nan Yan, Lan Wang

D2-1430_H3.7 Outlier Removal in MEG Data for Imagined Speech Classification 1435
Koki Nose, Hajime Yano, Tetsuya Takiguchi, Seiji Nakagawa

D2-1430_H3.8 Performance Evaluation of CHIRPS and ETCCDI Indices for Extreme Rainfall Risk Mapping in Thailand Using XGBoost 1441
Vinitar Khettar, Nuntikorn Kitratporn, Sawarin Lerk-u-suke, Jirabhorn Chaiwongsai, Phaisarn Jeefoo, Chanika Sukawattanavijit

D2-1430_H3.9 Riverbed Estimation Using Locally-Structured Unitary Network 1446
Seiyu Hitomi, Hiroyasu Yasuda, Kiyoshi Hayasaka, Shogo Muramatsu

D2-1430_H3.10 Contrastive Learning of Temporal and Event-Based Behavioral Views for Universal User Embeddings 1452
Yuuki Tachioka

D2-1430_H3.11 Market Forecasting Using LSTM-ARIMA Model with MACD Decomposition 1458
Teng-Chih Yu, Jian-Jiun Ding

Multimodal AI

- D2-0830_P1.1** VoxRep: Enhancing 3D Spatial Understanding in 2D Vision-Language Models via Voxel Representation 1464
Alan (Gia Tuan Dao) Dao, Norapat Buppodom
- D2-0830_P1.2** Active Multi-Object Tracking for 3D Reconstruction with Hierarchical Reinforcement Learning 1470
Heng Li, Cheng Cai
- D2-0830_P1.3** Multimodal Sentiment Analysis with Missing Modality: A Knowledge-Transfer Approach 1476
Weide Liu, Huijing Zhan
- D2-0830_P1.4** Modeling Spatiotemporal Multimodal Data With Kernel Graph Regression Models And Copulas 1481
Jeffrey Wu, Gareth Peters
- D2-0830_P1.5** CopeCap: A lightweight image captioning model with collaborative prompt learning 1487
Xiwei Yu, Guoshun He, Huijing Zhan
- D2-0830_P1.6** Lyric-Aware Karaoke Background Video Selection Using Large Language Models and Moment Retrieval 1497
Tomoki Ariga, Jun Taniguchi, Yosuke Higuchi, Sayaka Toma, Kunihiro Abe, Rie Shigyo, Tetsuji Ogawa
- D2-0830_P1.7** Audio-Visual Speech Recognition Based on Cross-Lingual Transfer Learning 1503
Fumiya Kondo, Tamura Satoshi
- D2-0830_P1.8** Exploring Machine Learning and Language Models for Multimodal Depression Detection 1509
Javier Si Zhao Hong, Timothy Zoe Delaya, Sherwyn Chan Yin Kit, Pai Chet Ng, Xiaoxiao Miao

Signal & Information Processing I

- D2-1430_P2.1** Generalized Student's t Sparse Kernel Learning for Robust Signal Processing 1515
Long Pan, Libiao Peng, Xifeng Li, Dongjie Bi, Yongle Xie
- D2-1430_P2.2** A Hierarchical Attention Model for Local and Global Feature Integration in RCS Classification 1521
Yida Wu, Caiyun Wang, Jianing Wang, Xiaofei Li, Ying Nan

D2-1430_P2.3 A Sliding-Window Range–Bearing Scan STAP for Underwater Active Sonar Target Detection	1527
<i>Weisi Hua, Yixin Yang, Yuxuan Chen, Xianghao Hou</i>	
D2-1430_P2.4 TH-LDV: Transformer-based Hybrid method for Signal Detection in Laser Doppler Velocimetry	1532
<i>Yue Wang, Ruifeng Li, Changsong Liu, Liangrui Peng, Ning Ding, Gang Yao</i>	
D2-1430_P2.5 Estimating Dynamic Graph Flows with Kernel Models and Hadamard-Structured Riemannian Constraints	1538
<i>Duc Thien Nguyen, Konstantinos Slavakis, Dimitris Pados</i>	
D2-1430_P2.6 Period Estimation for Time-Varying Graph Signals and Its Application to Graph Wiener Filter	1544
<i>Tsutahiro Fukuhara, Junya Hara, Hiroshi Higashi, Yuichi Tanaka</i>	
D2-1430_P2.7 Computationally Efficient Sparse Signal Recovery by Deep Unfolded-Periodic Sketched ISTA	1550
<i>Tatsuki Tokumura, Ayano Nakai-Kasai, Tadashi Wadayama</i>	
D2-1430_P2.8 Fisher Information-based Metrics for Representation Learning	1556
<i>Do Nguyen Dang Thi, Le Quoc Anh, Tran Trong Duy, Le Vu Ha, Nguyen Linh Trung</i>	
D2-1430_P2.9 Wave Direction Estimation Based on Local Gradient Techniques from Satellite Imagery for Coastal Dynamics Monitoring	1562
<i>Woramet Simrum, Paweena Kanokhong, Chakapat Chokchaisiri, Somrudee Deepaisarn, Kittipisut Chansri, Chanyut Lisawat, Waranrach Viriyavit, Akkharawoot Takhom, Phutphalla Kong, Didin Agustian Permadi, Sharifah Hafizah Syed Ariffin, Surasak Boonkla, Kasorn Galajit, Jessada Karnjana</i>	
D2-1430_P2.10 HIQA-DB: A Benchmark Dataset for Image Quality Assessment in Hospital Surveillance	1568
<i>Yujin Han, Taewan Kim</i>	
D2-1430_P2.11 Semantic Neural View Synthesis for Key Content Preservation in Horizontal-to-Vertical Video Conversion	1572
<i>Dipanita Chakraborty, Minoru Okada, Kosin Chamnongthai</i>	
Signal & Information Processing II	
D3-1030_P1.1 Low-Complexity Total Variation-based Signal Reconstruction with Adaptive Gradient Descent for Compressive sensing	1578
<i>Pei-Cheng Yeh, Chieh-Li Wang, Yuan-Hao Huang</i>	
D3-1030_P1.2 Robust Initialization Strategies for Hankel Structured Low-Rank Approximation via Variable Projection	1584
<i>Natsuki Yoshino, Akira Tanaka</i>	

D3-1030_P1.3 High-Resolution ISAR Imaging for High-Speed Targets via Joint Intra-Pulse and Inter-Pulse Translational Motion Compensation	1590
<i>Jiabao Wang, Shuai Shao, Jiaqi Wei</i>	
D3-1030_P1.4 Sparse Echo Reconstruction of Micro-motion Targets Under the Joint Constraints of Low-rank and Periodic Consistency	1596
<i>Mingming Jin, Jun Wang, Shaoming Wei, Peng Lei</i>	
D3-1030_P1.5 Distributed Extended Object Tracking with Adaptive Networks	1602
<i>Kaidi Yang, Wei Xia</i>	
D3-1030_P1.6 Extended Object Tracking: A DNN-aided approach	1608
<i>Runhe Gan, Wei Xia</i>	
D3-1030_P1.7 Non-negative Learned ISTA with Reflected-ReLU-Augmented L1 Regularization	1615
<i>Haruki Esaki, Towa Yasui, Seisuke Kyochi</i>	
D3-1030_P1.8 Phoneme-Specific Challenges to Intelligibility in Hearing Impairment Under Noisy Condition	1621
<i>Denawati Junia, Candy Olivia Mawalim</i>	
Biomedical Signal Processing and Systems I	
D2-0830_P2.1 Predicting Problematic Internet Use in Children Using Feature-Rich Structured Data with Ensemble Machine Learning and Bayesian Optimisation	1627
<i>Niteesh K R, Pooja T S</i>	
D2-0830_P2.2 Phonocardiogram Signal Analysis for Myocardial Infarction Level Prediction Using Deep Learning Model	1633
<i>Ira Puspasari, Tati L.R. Mengko, Agung W. Setiawan, Miftah Pramudyo, Nobuo Watanabe, Trio Adiono</i>	
D2-0830_P2.3 Prediction of Maximum and Minimum Postprandial Blood Glucose Levels in People with Diabetes	1639
<i>Kotaro Nagayama, Shota Kato, Kana Eguchi, Masahide Hamaguchi, Hiroyuki Tominaga, Youji Hamaguchi, Michiaki Fukui, Manabu Kano</i>	
D2-0830_P2.4 Towards Telepathic Communication: A Multi-Band EEG Model for Imaginary Speech Decoding	1645
<i>Yifan Zhang, Yuting Ding, Fei Chen</i>	
D2-0830_P2.5 Tiny-VRN: A Lightweight Variational Residual Network for EEG-Based Emotion Recognition	1651
<i>Sivaraj Nimishan, Selvarajah Thuseethan, Shanmuganathan Vasanthapriyan, Roshan G. Ragel</i>	

D2-0830_P2.6 A Comparison of Solicited and Longitudinal Cough Sounds for Tuberculosis Detection	1657
<i>Aprianto Dwi Prasetyo, Bagus Tris Atmaja, Dhany Arifianto, Sakriani Sakti</i>	
D2-0830_P2.7 Detecting Defecation Premonition from the Acoustic Activity of Bowel Sounds	1663
<i>Shota Miyagawa, Toshitaka Yamakawa, Masayuki Tanabe, Kazushi Ikeda</i>	
D2-0830_P2.8 EegCNR: A Novel Feature for Attention Estimation from EEG	1669
<i>Asif M S, Sagila Gangadharan K, Achutavarrier Prasad Vinod</i>	
D2-0830_P2.9 Lower Limb Calf Muscle Segmentation from Diffusion-Weighted Magnetic Resonance Images Using Deep Learning	1675
<i>Eshan Pandey, Xiaomeng Wang, Julian Gan, Ying-Hwey Nai, Derek Hausenloy, Pek Lan Khong, Forest Su Lim Tan, Thiruneepan Selvakulasingam, Ryan Fraser Kirwan, Cheryl Pei Ling Lian</i>	
D2-0830_P2.10 Principal Component Regularization in Iterative Inversion of DBIM for Ultrasound Tomography	1681
<i>Nguyen Thi Thu, Tran Quang-Huy, Luong Thi Theu, Duc-Tan Tran</i>	
D2-0830_P2.11 Reasoning Visualization for Critical Care EEG Classification with Prototypical Part Networks	1688
<i>Takuma Bingo, Hajime Yano, Taichiro Ashizaki, Kazuma Koda, Masaya Togo, Riki Matsumoto, Tetsuya Takiguchi</i>	
D2-0830_P2.12 Plant Species-Specific Anomaly Detection Based on Electrophysiological Signals	1694
<i>Andy Desman Lo, Elvin Nur Furqon, Junaidul Islam, Isack Farady, Kahlil Muchtar, Ronnie Concepcion II, Chih-Yang Lin</i>	
Biomedical Signal Processing and Systems II	
D2-1030_P1.1 Freeze and Learn using KAN for Infant Cry Classification	1700
<i>Arth Shah, Vishnu Vardhan, Hemant Patil</i>	
D2-1030_P1.2 Investigation of Enhancement Strategies for Recurrent Spiking Neural Network based Brain-Machine Interface Decoding	1706
<i>Wilson Tansil, Nur Ahmadi, Timothy Constandinou, Dessi Puji Lestari</i>	
D2-1030_P1.3 Detecting Deceptive Responses Due to Psychological Bias by the Probability Density Function of EEG Content Rate Dynamics During NEO-FFI Answering	1712
<i>Yuto Ashikawa, Yosuke Kurihara</i>	
D2-1030_P1.4 A Comparative Analysis of Statistical, Regional CNN, and Sequential Transformer Approaches for Alzheimer's Disease Classification	1718
<i>Trí Huynh, Xuan Hoc Pham, Nhu Nguyen, Thi Thu Nguyen, Huong Ha, Lua Ngo</i>	

D2-1030_P1.5 Beyond Speech and More: Investigating the Emergent Ability of Speech Pre-Trained Models for Classifying Physiological Time-Series Signals 1724
Orchid Chetia Phukan, Swarup Ranjan Behera, Girish, Mohd Mujtaba Akhtar, Arun Balaji Buduru, Rajesh Sharma

D2-1030_P1.6 Channel Selection Guided by Layer-wise Relevance Propagation for CNN-Based EEG Classification of Major Depressive Disorder 1730
Woo-Seok Ahn, Seung-Hwan Lee, Han-Jeong Hwang

D2-1030_P1.7 Development of HRV-Based Biomarkers for Predicting Blood Glucose Levels 1734
Ju-An Park, Jun-Seok Lee, Na-Ri Kim, Han-Jeong Hwang

D2-1030_P1.8 Development of 3D Textile Electrodes for Electrocardiography Measurement 1738
Sang-Ho Lee, In-Su Park, Han-Jeong Hwang

Advanced Signal Processing and Resource Management for Sensing and Communication

D3-0830_P1.1 A reinforcement learning-based approach to cooperative multi-UAV task allocation 1742
Naohiro Kubota, Hideyoshi Miura, Tomotaka Kimura, Kouji Hirata

D3-0830_P1.2 Trajectory Design of UAVs-Assisted Edge Computing Systems for Efficient Data Collection from Animal Herds 1746
Nao Maeda, Tomotaka Kimura, Kouji Hirata

D3-0830_P1.3 Priority-based RCSA method considering required frequency slot width in multi-core fiber networks 1750
Funa Fukui, Yutaka Fukuchi, Kouji Hirata

D3-0830_P1.4 Retraining-Free Blockage Prediction for Millimeter-Wave Communications Based on Minor Components of Angular Power Profiles 1755
Xiaoqing Tong, Kohei Mitani, Kazunori Hayashi, Koji Yamamoto, Takuto Arai, Shuki Wai, Tatsuhiko Iwakuni, Daisei Uchida

D3-0830_P1.5 Modified Resource Allocation Algorithm based on Co-channel Interference Prediction in Local 5G Environmentts 1761
Takeru Nanjo, Osamu Takyu

D3-0830_P1.6 Implicit Interference Status Notification Through Time & Frequency Resource Selection in LoRaWAN 1767
Yuto Hayasaka, Koichi Adachi

D3-0830_P1.7 Wireless Environment Estimation with Directional Antennas using Radio Environment Database for Wireless Information and Power Transfer in Smart Factories 1772
Kohei Yuzawa, Zhengdong Lin, Yu Kagaya, Yoshiaki Narusue, Takeo Fujii

D3-0830_P1.8 Data-Driven Tuning of Neural Network Aided Least Squares for UWB-TDoA Indoor Positioning 1778
Ryoichi Kawaguchi, Shinsuke Ibi, Hisato Iwai

Wireless Communications & Networking

D1-1430_P2.1 Low-Complexity Sparse Channel Estimation for Reconfigurable Intelligent Surface-Aided MIMO 1784
Wei-Lin Chiang, Shu-Yu Lin, Jung-Chun Chi, Yuan-Hao Huang

D1-1430_P2.2 BFIS: Efficient Unknown Protocol Feature Extraction Method For Satellite Communication Systems 1790
Xianwen Ling, Kun Zhang, Rong Tong, Dianying Chen

D1-1430_P2.3 Outdoor Experiment of Deep Joint Source-Channel Coding Using FFT-Enabled Convolutional Neural Network for Image Transmission 1796
Tomoka Mori, Hiroshi Tatsukawa, Yuji Kawai, Yoshinori Shinohara, Hiroki Ikeda, Daisuke Hisano

D1-1430_P2.4 On LSTM-Based Behavioral Modeling of Radio-Frequency Power Amplifiers with a Small Training Dataset 1801
Ryoki Yamaguchi, Satoshi Miyata, Suehiro Shimauchi, Eiji Mochida, Seiji Fujiwara

D1-1430_P2.5 DL-based Optical Fibre Fault Detection for Healthcare Telesurgery Communication System 1806
Khushi Shah, Lakshit Pathak, Akshita Abrol, Kanak Jain, Rajesh Gupta, Parish Shah, Sudeep Tanwar, Umesh Bodkhe, Tong Rong

D1-1430_P2.6 Overcoming Imperfect Detection Limitations: Deep Learning-Based Calibration Strategy for Rotating Interferometric Arrays 1812
Zhaohang Zhang, Chunzhe Wang, Zhen Huang, Yafeng Zhan

D1-1430_P2.7 A Regional Clustering Method Based on Propagation Similarity for Modeling Cumulative Interference from Large Numbers of Terminals 1818
Tatsuro Hidaka, Osamu Takyu, Kei Inage, Takeo Fujii, Kohei Yoshida, Masayuki Ariyoshi

D1-1430_P2.8 Radio Frequency Fingerprinting-Based Device Identification Using Deep Metric Learning 1824
Dinh Tuan Anh, Bui Tung Lam, Pham An Duy, Pham Minh Tuan, Tran Vinh Co, Nguyen Huu Tinh, Huynh Cong Bang

D1-1430_P2.9 GNSS Spoofing Detection Based on LSTM-TNN-CVAE Network 1830
Chaowen Tang, Tian Qin

D1-1430_P2.10 Enhancing Speech Quality in Scintillating Satellite Communications: A Rician Fading Modeling Approach 1835
Teh Kah Kuan, Sun Hanwu, Tran Huy Dat

Advanced Multimedia Applications

D1-1630_P2.1 Ensemble Methods for Estimating the Localization of Coronary Stenosis from CT Images Using 3D CNN Models	1841
<i>Minori Kondo, Masaki Aono, Kazuki Shimizu, Masashi Hashimoto, Takeshi Miyaji, Kei Nomura</i>	
D1-1630_P2.2 Tiered Assessment for DSP Education: Exploring Students' Motivation and Performance	1847
<i>Eliathamby Ambikairajah, Tharmakulasingam Sirojan, Vidhyasaharan Sethu</i>	
D1-1630_P2.3 An Investigation of Parameter Scheduling for Image Restoration in Optical Analog Circuits	1853
<i>Taisei Kato, Ryo Hayakawa, Soma Furusawa, Kazunori Hayashi, Youji Iiguni</i>	
D1-1630_P2.4 Robust Cloud Removal from Optical Satellite Images Using Synthetic Aperture Radar and Multimodal Embedding Prior	1859
<i>Taishin Miura, Shunsuke Ono, Ryo Matsuoka</i>	
D1-1630_P2.5 Reflection and Noise Separation from Polarized Images via Joint Nonnegative Matrix Factorization and Plug-and-Play Denoising	1864
<i>Maharu Oda, Ryo Matsuoka</i>	
D1-1630_P2.6 Gated probabilistic diffusion for temporal action segmentation	1868
<i>Yun Li, Hanmin Li, Kin-Man Lam</i>	
D1-1630_P2.7 Theory of Spherical VR model for Landscape Representation	1874
<i>Hiroyuki Nishimoto, Toru Takahashi, Masakazu Yoshida</i>	
D1-1630_P2.8 HyTver: A Novel Loss Function for Longitudinal Multiple Sclerosis Lesion Segmentation.	1880
<i>Dayan Perera, Ting Fung Fung, Vishnu Monn Baskaran</i>	
D1-1630_P2.9 KH-FUNSD: A Hierarchical and Fine-Grained Layout Analysis Dataset for Low-Resource Khmer Business Document	1886
<i>Nimol Thuon, Jun Du</i>	
D1-1630_P2.10 Effective Speckle Noise Reduction Using Transformed Bayesian Likelihood with Wiener-Based and Sketch-Based Geometric Priors	1892
<i>Ming-Hsun Mo, Pin-Wen Huang, Jian-Jiun Ding</i>	

Advances in Multimodal AI for Multimedia Applications

D2-1630_P2.1 Efficient Generative Adversarial Networks for Color Document Image Enhancement and Binarization Using Multi-scale Feature Extraction	1898
<i>Rui-Yang Ju, KokSheik Wong, Jen-Shiun Chiang</i>	

D2-1630_P2.2 Leveraging Large Language Models in Visual Speech Recognition: Model Scaling, Context-Aware Decoding, and Iterative Polishing	1904
<i>Zehua Liu, Xiaolou Li, Li Guo, Lantian Li, Dong Wang</i>	
D2-1630_P2.3 Computationally-efficient Call Classification of New Zealand Birds using Texture-based Features	1910
<i>Yonghui Tao, Mathis Quere, Yusuke Hioka, Stephen Marsland</i>	
D2-1630_P2.4 Incorporating Semantic Visual Content into Click-Through Rate Prediction for Video Advertisements	1916
<i>Yoshiaki Tanabe, Shuntaro Masuda, Gakumatsu Ryu, Naoto Tanji, Hiroyuki Seshime, Ling Xiao, Toshihiko Yamasaki</i>	
D2-1630_P2.5 From Blurry to Brilliant Detection: YOLO-Based Aerial Object Detection with Super Resolution	1922
<i>Ragib Amin Nihal, Benjamin Yen, Takeshi Ashizawa, Katsutoshi Itoyama, Kazuhiro Nakadai</i>	
D2-1630_P2.6 ATJO: Adaptive three-dimensional joint optimization for remote sensing video super-resolution	1928
<i>Tian Qin, Lijing Bu, Zhengpeng Zhang, Mingjun Deng, Yin Yang, Jingxue Wang, Xinyu Lan, Wenjuan Peng, Yang Hu</i>	
D2-1630_P2.7 Block-level Lagrange multiplier adaptation based on distortion propagation factors	1934
<i>Hongwei Guo, Yipeng Liu, Lei Luo, Chengbiao Fu, Ce Zhu</i>	
D2-1630_P2.8 Distributed Compressed Video Sensing with Enhanced Boundary Handling Based on Extended Convolutional Sparse Representation	1940
<i>Ibuki Muta, Yoshimitsu Kuroki</i>	
D2-1630_P2.9 Joint Modeling of Big Five and HEXACO for Multimodal Apparent Personality-trait Recognition	1946
<i>Ryo Masumura, Shota Orihashi, Mana Ichori, Tomohiro Tanaka, Naoki Makishima, Taiga Yamane, Naotaka Kawata, Satoshi Suzuki, Taichi Katayama</i>	
D2-1630_P2.10 Foreground-Background Segmentation Based Surveillance Video Coding	1952
<i>Jiyong Yu, Luheng Jia, Yifan Zang, Zhaoyang Yu, Shuyuan Zhu, Li Song, Kebin Jia</i>	
D2-1630_P2.11 Rain Removal via VAE-Enhanced Transformer with Hierarchical Feature Integration	1958
<i>Yaya Huang, Litong Liu, KokSheik Wong</i>	

Scalable and Efficient Signal Processing for Multimodal AI Systems (ESPRESSO)

D3-1430_P2.1 Exploring Audio-Visual Fusion Methods in Foundation Model-Based Deception Detection 1964
Jiaxiang Meng, Hardik B. Sailor, Qiongqiong Wang, Tianchi Liu, Kong Aik Lee, Xingmei Wang

D3-1430_P2.2 Emot-CM-BERT: Adaptive Attention and Class-Aware Cross-Modal Learning for Emotion Recognition from Audio and Text 1969
Shintami Chusnul Hidayati, James Rafferty Lee, Kevin Davi Samuel

D3-1430_P2.3 DP-GS: Depth-prior & Perception-guided Gaussian Splatting for Sparse-view Novel View Synthesis 1975
Bowen Gao, Zhicheng Lu, Mingyi He, Yuchao Dai

D3-1430_P2.4 Efficient Video to Audio Mapper with Visual Scene Detection 1981
Mingjing Yi, Yuxi Wang, Ming Li

D3-1430_P2.5 Adversarial Learning for Duration Prediction in Indonesian Text-to-Speech: Modification to Stochastic and Deterministic Predictors 1986
Yoga Tiara Wiguna, Bima Prihasto, Boby Mugi Pratama, Chia-Hung Yeh, Jia-Ching Wang

D3-1430_P2.6 Narrativity-Aware Video Summarization Based on Vision and Language Foundation Models 1991
Shumpei Saito, Hiroyuki Ueda, Yosuke Ito, Kazuyoshi Yoshii

D3-1430_P2.7 RawTFNet: A Lightweight CNN Architecture for Speech Anti-spoofing ... 1997
Yang Xiao, Ting Dang, Rohan Kumar Das

D3-1430_P2.8 Dynamic Fusion Multimodal Network for SpeechWellness Detection 2002
Wenqiang Sun, Han Yin, Jisheng Bai, Jianfeng Chen

D3-1430_P2.9 AIGuard: Anomaly Detection in Surveillance Videos with YOLOv8 2008
Rungpilin Anantathanavit, Supakorn Suthirat, Po-Chyi Su

D3-1430_P2.10 Ensemble Confidence Calibration for Sound Event Detection in Open-environment 2014
Yuanjian Chen, Han Yin

D3-1430_P2.11 Enhancing Stereo Sound Event Detection with BiMamba and Pretrained PSELDnet 2020
Wenmiao Gao, Han Yin

Multimedia Security & Forensics

D1-1430_P1.1 The Potential of LLMs for Generating Malicious Domain Names 2026
Lim Kit Michael Ye, Kaijian Zheng, Ngai Fong Law, Jianping Li

D1-1430_P1.2 Reducing Implicit Class Imbalance in Unlabeled Datasets Using Text-Specified Sensitive Attributes 2032
Kosei Suyama, Kazuaki Nakamura

D1-1430_P1.3 DRASP: A Dual-Resolution Attentive Statistics Pooling Framework for Automatic MOS Prediction 2038
Cheng-Yeh Yang, Kuan-Tang Huang, Chien-Chun Wang, Hung-Shin Lee, Hsin-Min Wang, Berlin Chen

D1-1430_P1.4 Multimodal Large Language Model for Deepfake Video Detection and Description 2044
Haoran Sun, Chen Cai, Kong Aik Lee, Lap-Pui Chau, Yi Wang

D1-1430_P1.5 Biometric Identification Using Default Mode Network Features Extracted from Eyes-Open Resting-State EEG Data 2050
Parvathy Remesh, Jijomon Chettuthara Moncy, Vinod Achutavarrier Prasad

D1-1430_P1.6 Backdoor Poisoning Attack Against Face Spoofing Attack Detection Methods 2056
Shota Iwamatsu, Koichi Ito, Takafumi Aoki

D1-1430_P1.7 Access Control for Diffusion Models by Random Masking the Covariance of Initial Noise Distribution 2062
Temma Tanaka, Kazuaki Nakamura

Privacy and Security in Speech AI

D3-1430_H3.1 Voice Privacy Protection with Adversarial Examples Using Anchor Speaker Embedding 2068
Shunya Ishikawa, Yuki Katsumata, Toru Nakashika

D3-1430_H3.2 Investigation of perception inconsistency in speaker embedding for asynchronous voice anonymization 2074
Rui Wang, Liping Chen, Kong Aik Lee, Zhengpeng Zha, Zhenhua Ling

D3-1430_H3.3 SegReConcat: A Data Augmentation Method for Voice Anonymization Attack 2080
Ridwan Arefeen, Xiaoxiao Miao, Rong Tong, Aik Beng Ng, Simon See

D3-1430_H3.4 An Enhanced Probabilistic Approach for Singfake Generation 2086
Arth Shah, Aniket Pandey, Satyam R. Tiwari, Hemant Patil

D3-1430_H3.5 Neural Semi-fragile Watermarking for Proactive Deepfake Speech Detection 2092
Dohyun Yoon, Tomoki Toda

D3-1430_H3.6 Investigating Self-supervised Learning-Based Front-End for Multi-Channel Replay Attack Detection 2098
Takuo Yamaguchi, Sayaka Shiota, Naohiro Tawara

D3-1430_H3.7 Transferability of Adversarial Examples across Speaker Embedding Models for Voice Privacy Protection 2104
Kotaro Nakamura, Takuya Takahashi, Toru Nakashika

D3-1430_H3.8 Voice Privacy Preservation with Multiple Random Orthogonal Secret Keys: Attack Resistance Analysis 2110
Kohei Tanaka, Hitoshi Kiya, Sayaka Shiota

D3-1430_H3.9 CycleSiFiNF-VC: Controllable Non-Parallel Voice Conversion by Neural Formant Manipulation with Improved Cycle-Consistency Loss 2116
Sumiharu Kobayashi, Takashi Nose, Akinori Ito

D3-1430_H3.10 Recoverable Audio Adversarial Examples for Voice Protection in One-shot Voice Conversion 2122
Chenshuai Shu, Tianpeng Zheng, Yanxiang Chen

D3-1430_H3.11 Reference-free Adversarial Sex Obfuscation in Speech 2128
Yangyang Qu, Michele Panariello, Massimiliano Todisco, Nicholas Evans

Recent Advances in Multimedia Enrichment, Security and Privacy

D2-1630_H3.1 Reversible Data Hiding in EtC Images with Flexible Access Privileges 2134
Yusaku Kato, Shoko Imaizumi

D2-1630_H3.2 Robust Ownership Verification of DNN Models Against JPEG Compression via Probability-Controlled Adversarial Attacks 2140
Teruki Sano, Minoru Kuribayashi, Masao Sakai, Shuji Ishobe, Eisuke Koizumi, Zhang Zhang

D2-1630_H3.3 Detoxification of Poisoned Recognition Models by Fine-tuning with Out-of-Distribution Samples 2146
Junsuke Takano, Kazuaki Nakamura

D2-1630_H3.4 Layer-Wise Weight Statistics for Node Classification and Defense of Federated Large Language Models 2152
Alexander Berns, Reon Akai, Minoru Kuribayashi, Rémi Cogranne

D2-1630_H3.5 Robustness evaluation against fine-tuning in associative watermarking method for CNN 2158
Keiichi Mori, Masaki Kawamura

D2-1630_H3.6 Lossless Image Processing for OpenEXR Images with Flexible Functions 2164
Anna Yamaguchi, Shoko Imaizumi

D2-1630_H3.7 Proposal of a Random Encoding Layer Compatible with Arbitrary Message Lengths for DiffuseTrace 2170
Ou Egami, Masaki Kawamura

D2-1630_H3.8 Automatic Dependent Surveillance-Broadcast Preamble Classification for Spoofing Detection 2176
Darren Kah Hou Quek, Guang Hua, Zhiping Lin

D2-1630_H3.9 Model Extraction Attack and Its Countermeasure for Denoising Diffusion Implicit Models 2181
Hayato Shoji, Kazuaki Nakamura

D2-1630_H3.10 Content-Aware Dominant Color Extraction and Its Application to Multiple-key-Color Image Retrieval 2187
Mei Hashimoto, Michiharu Niimi

D2-1630_H3.11 Privacy-Preserving Image Retrieval Scheme Using Combined Features in Cloud Computing 2193
Jing Liang, Yuxuan Wang, Tingting Song, Ce Zheng, Peiya Li

Tracing the Fake: Deepfake Detection, Attribution & Spoof-aware Speaker Verification Across Languages & Accents

D1-1630_L1.1 NE-PADD: Leveraging Named Entity Knowledge for Robust Partial Audio Deepfake Detection via Attention Aggregation 2199
Huhong Xian, Rui Liu, Berrak Sisman, Haizhou Li

D1-1630_L1.2 Robust Localization of Partially Fake Speech: Metrics and Out-of-Domain Evaluation 2205
Hieu-Thi Luong, Inbal Rimon, Haim Permuter, Kong Aik Lee, Eng Siong Chng

D1-1630_L1.3 Mixture of Low-Rank Adapter Experts in Generalizable Audio Deepfake Detection 2211
Janne Laakkonen, Ivan Kukanov, Ville Hautamäki

D1-1630_L1.4 Continual Audio Deepfake Detection via Universal Adversarial Perturbation 2217
Wangjie Li, Lin Li, Qingyang Hong

D1-1630_L1.5 Exploring Source Features with Deep Residual Neural Networks for Replay Attack Detection 2223
Suresh Veesa, Badugu Vamsi Krishna, Madhusudan Singh

D1-1630_L1.6 A Preliminary Study on Sectional Voice Anonymization and Detection ... 2229
Shaoqi Tang, Zeyan Liu, Liping Chen, Kong Aik Lee, Tomoki Toda, Zhenhua Ling

D1-1630_L1.7 ArcticEcho: A Novel Speaker-Controlled Voice Cloning Dataset for Modern Deepfake Detection Benchmarking 2235
Soham Gangopadhyay, Inderpreet Singh, Prateek Pandya, Ashish Mani, Sumit Goswami

D1-1630_L1.8 Variational Regularization for End-to-end Speech Deepfake Detection ... 2241
Siqing Qin, Kong Aik Lee, Man-Wai Mak, Pasquale Lisena, Massimiliano Todisco

D1-1630_L1.9 A Wavelet tour of Audio Deepfake Detection 2247
Arth Shah, Aniket Pandey, Manav Giakwad, Hemant Patil

D1-1630_L1.10 Fusion of Modulation Spectrogram and SSL with Multi-head Attention for Fake Speech Detection 2253
Rishith Sadashiv T N, Abhishek Bedge, Saisha Suresh Bore, Jagabandhu Mishra, Mrinmoy Bhattacharjee, S R Mahadeva Prasanna

Emerging Technologies and Applications of Image Processing and Computer Vision

D2-1030_H3.1 You Only Touch Once: One-Touch System for Personalized 3D Music Video Generation 2259
Kyungjune Lee, Youngjin Shin, Jungwoo Huh, Sanghoon Lee

D2-1030_H3.2 Single-Image Pupil Localization via Implicit 3D Eye Reconstruction 2264
Taejun Roh, Yejin Cho, Duong Hai Nguyen, Chul Lee

D2-1030_H3.3 Flow-Guided Consistent Video Depth Estimation for Cross-Dataset Generalization 2270
Jaeseok Jang, Chang-Su Kim

D2-1030_H3.4 DCB: An Efficient Approach for Building Long-Range Dependencies in CNNs 2276
Tianxiang Lan, Mingyi He, Yuchao Dai

D2-1030_H3.5 A User-Guided and Local Motion-Adaptive Framework for Virtual Product Placement in Video 2282
Tianwen Zhang, Ju-Won Seo, Kang-Min Kim, Keunsoo Ko

D2-1030_H3.6 Shallow yet Perceptual Decoding for Neural Image Compression through Minimal Nonlinearity 2287
JaeKyung Ryu, Nam Ik Cho

D2-1030_H3.7 SyncScore: A Framework for Synchronization Scoring in Group Sports via Human Pose Estimation 2293
Khai Pin Ang, Iven Zi Yin Low, Yumun Hooj, Yuen Peng Loh

D2-1030_H3.8 Data Augmentation-Driven Segmentation of Ovarian Tumor Ultrasound Images using Vision Mamba 2299
Thanh-Phuc Dao, Huyen-Trang To, Hoang-Son Bui, Thi-Lan Le

D2-1030_H3.9 Optimizing JPEG Decoder for Bitstream-Corrupted Image Restoration .. 2305
Shumin Jiang, Hao Qin, Tianyi Liu, Yi Wang

D2-1030_H3.10 Semantic Scene Completion from a Single Depth Image with Coarse-Grained Segmentation 2311
Jiun Yen Ching, Lai-Kuan Wong, Wai Lee Kung

D2-1030_H3.11 Pixel-weighted Domain Adaptation for Agricultural Segmentation 2317
Shunta Kimura, Handie Shao, Shogo Matsumoto, Daiki Yamada, Toshihiro Kitajima, Hideki Nakayama

Generative AI Models for Vision-Based Applications

D1-1430_H3.1 TRUST: Token-dRiven Ultrasound Style Transfer for Cross-Device Adaptation 2323
Nhat-Tuong Do-Tran, Ngoc-Hoang-Lam Le, Ian Chiu, Po-Tsun Paul Kuo, Ching-Chun Huang

D1-1430_H3.2 Two-Stage Transformer-based Deep Hyperspectral and Multispectral Image Fusion Network for Hyperspectral Image Super-Resolution 2330
Wo-Yen Li, Chia-Ming Lee, Chih-Chung Hsu, Volodymyr Khylenko, Li-Wei Kang

D1-1430_H3.3 Pedestrian Detection based on Visible Guided Occlusion Handling 2336
Lien-Chieh Huang, Ching-Te Chiu, Yung-Cheng Su

D1-1430_H3.4 Spatial-Frequency Guided Moiré Removal with Multi-Stage Feature Fusion 2342
Chen Lo, Chia-Hung Yeh

D1-1430_H3.5 Registration of Infrared and Visible Images Using Style Transfer-Based Semantic Segmentation 2347
Si-Ting Lin, Chih-Hung Han, Chieh-Ling Lee, Po-Chyi Su, Feng-Tsun Chien, Min-Kuan Chang

D1-1430_H3.6 Peransformer: Improving Low-informed Expressive Performance Rendering with Score-aware Discriminator 2353
Xian He, Wei Zeng, Ye Wang

D1-1430_H3.7 Prompt-Based Vertebral Segmentation Using a Generative AI Approach in OVCF Spinal Radiographs 2359
Po-Kai Su, Pei-Rong Jiang, Kai-Xuan Xu, Meng-Lei Su, Jiann-Her Lin, Hsin-Han Chiang, Hsiao-Chi Li

D1-1430_H3.8 A Dual-Stream Diffusion Model with Physically-Based Rendering for Single Image Reflection Removal 2365
Cheng-Wei Hsu, Ming-Sui Lee

D1-1430_H3.9 DYNAMIC FACIAL EXPRESSION RECOGNITION IN THE WILD USING MAMBA-STYLE SELECTIVE SSM AND FACIAL ATTENTION MECHANISM 2371
Yudhistira Arditya Pratama, Theophilus Ezra Nugroho Pandin, Yi-Zeng Hsieh

Recent Advances in Multimodal Learning for Computer Vision

D3-1030_H3.1 Allegory of the Cave: Breakdown of Illusions in Multimodal Perception with Neural Radiance Fields 2376
Axel Paivansalo, Ching-Chun Chang, Hanrui Wang, Futa Waseda, Isao Echizen

D3-1030_H3.2 Overlapped Coffee Beans Detection and Localization Using a Low-Cost 3D Monocular Point Cloud Clustering Method 2382
Isack Farady, Alifya Febriana, Chih-Yang Lin

D3-1030_H3.3 Interpretable Video-Text Alignment (VTA) for Cross-Modal Retrieval 2388
Tsung-Shan Yang, Yun-Cheng Wang, Chengwei Wei, Suyu You, C.-C. Jay Kuo

D3-1030_H3.4 Sequence Modeling and Generative Model Driven Non-Rigid 3D Reconstruction 2394
Yuxin He, Hui Deng, Mingyi He, Yuchao Dai

D3-1030_H3.5 Robust Audio-Visual Speech Recognition in Noisy Clinical Environments 2400
Akshita Abrol, Ridwan Arefeen, Haotong Yu, Alexi George, Kelvin Li, Zhengkui Wang, Rong Tong

D3-1030_H3.6 Integrating Visual XAI and LLMs for Interpretable Medical Image Analysis 2406
Chern Hong Lim, Xin Hui Lor

D3-1030_H3.7 InternVL-VPR: Hierarchical Zero-Shot Visual Place Recognition with VLM-driven Re-Ranking 2412
Zhi Hu, Liang Liao, Weisi Lin

D3-1030_H3.8 Token Compression Meets Compact Vision Transformers: A Survey and Comparative Evaluation for Edge AI 2418
Phat Nguyen, Ngai-Man Cheung

D3-1030_H3.9 Adapting Vision-Language Models for Information Extraction from Bilingual Medical Invoices 2424
Ha Do, Anh Dung DO, Thanh-Ha DO

D3-1030_H3.10 Zero-shot Artistic Text Recognition with Multimodal Language Models 2430
Tien Do, Thuyen Tran, Duy-Dinh Le, Thanh Duc Ngo

D3-1030_H3.11 Attention Based Deep Reference Frame Enhancement for VVC Inter Prediction 2436
Linchen Xu, Zhikai Liu, Fan Liang

Research Frontiers in Learned Visual Data Coding and Processing

D2-0830_H3.1 Neural Implicit Representations for Object-centric Machine Vision Tasks 2442
Yeoneui Kim, Je-Won Kang

D2-0830_H3.2 GoP-to-Frame Encoder Adaptation for Learned Video Compression 2448
Xiaohan Pan, Runsen Feng, Henan Wang, Yixin Gao, Zhibo Chen

D2-0830_H3.3 Efficient Adversarial Attack and Training on Learned Image Compression 2453
Jun Kurihara, Heming Sun

D2-0830_H3.4 Accelerating VVC Inter-Frame Coding: A Lightweight CNN for Fast QTMT Partitioning 2459
Jui-Chen Luo, Jiann-Jone Chen, Tien-Ying Kuo, Yi-Fan Wu, Zhang Kai-Jie

D2-0830_H3.5 Multimodal Speech Analysis for Early Detection of Mild Cognitive Impairment: A Scalable Approach 2465
Muhammad Bilal, Waleed Abdulla, Gary Cheung, Lynette Tippett, Reza Shahamiri

D2-0830_H3.6 Boundary-Enhanced Attention Network for Breast Mass Segmentation 2471
Rong Chen, Karungaru Stephen, Kenji Terada, Linhuang Wang

D2-0830_H3.7 Scale and Rotation Estimation of Similarity-Transformed Images via Cross-Correlation Maximization Based on Auxiliary Function Method 2477
Shinji Yamashita, Yuma Kinoshita, Hitoshi Kiya

D2-0830_H3.8 Strong Eye Closure Detection in Children with Profound Intellectual and Multiple Disabilities Using Robust Temporal Difference Features 2482
Kaito Kosaki, Teppei Nakano, Mari Wakabayashi, Tomomi Sato, Tetsuji Ogawa

D2-0830_H3.9 A Rate-Quality Model for Learned Video Coding 2488
Sang NguyenQuang, Cheng-Wei Chen, Xiem HoangVan, Wen-Hsiao Peng

D2-0830_H3.10 Low-Light RAW Image Enhancement with Additive Parameterization and State Space Model 2494
Shugo Yamashita, Masaaki Ikehara

D2-0830_H3.11 Synthesizing and Restoring Weather-corrupted Images with Conditional Diffusion Models 2499
Youngho Go, Sung-Hak Lee

Best Paper Finalists

D3-0830_IB.1 Integrating Semantic Knowledge for Enhanced Weakly-Supervised Group Activity Recognition 2505
Muhammad Adi Nugroho, Jinyoung Park, Yeeun Seong, Changick Kim

D3-0830_IB.2 Directed Graph Dynamic Mode Decomposition for Nonlinear State-Space Modeling	2511
<i>Hiromu Kanauchi, Ryuto Ito, Hiroyasu Yasuda, Masaaki Nagahara, Yuichi Tanaka, Shogo Muramatsu</i>	
D3-1030_IB.1 Digital-Optical Hybrid Computation for Deep Unfolding-Aided MIMO Signal Detection	2517
<i>Takumi Nishiyama, Lantian Wei, Tadashi Wadayama</i>	
D3-1030_IB.2 Uncolorable Examples: Preventing Unauthorized AI Colorization via Perception-Aware Chroma-Restrictive Perturbation	2523
<i>Yuki Nii, Futa Waseda, Ching-Chun Chang, Isao Echizen</i>	
D3-1030_IB.3 Voice Conversion Augmentation for Speaker Recognition on Defective Datasets	2529
<i>Ruijie Tao, Zhan Shi, Yidi Jiang, Tianchi Liu, Haizhou Li</i>	
D3-1030_IB.4 Explainable Disentanglement on Discrete Speech Representations for Noise-Robust ASR	2535
<i>Shreyas Gopal, Ashutosh Anshul, Haoyang Li, Yue Heng Yeo, Hexin Liu, Eng Siong Chng</i>	
D3-1030_IB.5 Rethinking Robust ASR Strategies: Can Textual In-Context Learning Improve Acoustic Robustness?	2541
<i>Benita Angela Titalim, Faisal Mehmood, Sakriani Sakti</i>	
D3-1030_IB.6 Noro: Noise-Robust One-shot Voice Conversion with Hidden Speaker Representation Learning	2547
<i>Haorui He, Yucheng Song, Yuancheng Wang, Haoyang Li, Xueyao Zhang, Li Wang, Gongping Huang, Eng Siong Chng, Zhizheng Wu</i>	
Perspective Panel Discussions	
D2-1430_IB.2 Advancing Speech Quality Assessment Through Scientific Challenges and Open-source Activities	2552
<i>Wen-Chin Huang</i>	
D2-1430_IB.1 Progress and Challenges in DNN-based Objective Quality Assessment of Synthesized Speech	2558
<i>Erica Cooper</i>	
D1-1630_IB.1 Foundation Models as Guardrails: LLM- and VLM-Based Approaches to Safety and Alignment	2564
<i>Huy Nguyen, Pride Kavumba, Tomoya Kurosawa, Koki Wataoka</i>	
D1-1430_IB.1 Speaker Privacy and Security in the Big Data Era: Protection and Defense against Deepfake	2570

Liping Chen, Kong Aik Lee, Zhen-Hua Ling, Xin Wang, Rohan Kumar Das, Tomoki Toda, Haizhou Li

D2-1430_IB.3 Non-Intrusive Intelligibility Prediction for Hearing Aids: Recent Advances, Trends, and Challenges 2576
Ryandhimas Zezario

D2-1430_IB.4 From Evaluation to Optimization: Neural Speech Assessment for Downstream Applications 2582
Yu Tsao

D3-1430_IB.1 Normalization through Fine-tuning: Understanding Wav2vec2.0 Embeddings for Phonetic Analysis 2587
Yiming Wang, Jiahong Yuan

D3-1430_IB.2 Enabling Internationalization of Affective Speech Technology using LLMs 2592
Bo-Hao Su, Shinji Watanabe, Chi-Chun Lee

Late Breaking Reports

D3-1430_P1.1 BiGaitNet: Deep CNN-Based Classification of Parkinson’s Disease Gait Abnormalities Using a Smart Insole Robust to Fewer Plantar Sensors 2598
Eun-Seo Park, Xianghong Liu, Chang-Hee Han

D3-1430_P1.2 Nonlinear System Identification Approach under Noisy Input Signals and Impulse Observed Noise by Kernel Adaptive Filtering Algorithm 2600
Ying-Ren Chien, En-Ting Lin

D3-1430_P1.3 Retinal Artery–Vein Segmentation via Attention-Guided W-Net and GAN-Based Boundary Refinement 2602
Jing-Ming Guo, De-Yu Guu, Yih-Ping Luh, Yi-Chong Zeng

D3-1430_P1.4 Local Contrast Enhancement in LDR Images via Adaptive Distribution of Clipped-histogram Excess 2604
Seong-hyun Jin, Dong-Min Son, Young-Ho Go, Sung-Hak Lee

D3-1430_P1.5 A Lightweight and Reversible Audio Watermarking Scheme Based on Integer Wavelet Transform 2606
Xuping HUANG, Akinori Ito

D3-1430_P1.6 Variance-driven U-Net Training and Chroma-scale-based Multi-exposure Image Fusion 2608
Changwoo Son, Youngho Go, Seunghwan Lee, Sunghak Lee

D3-1430_P1.7 Joint Design of Low Sidelobe Radar Waveform and Filter with Hardware Platform Verification 2610
Haoqian Rong, Shaojie Wang, Zining Zhao, Jiawei Zhang