

IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY



Association for
Computing Machinery

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine



2021

VOLUME 29

ITASFA

(ISSN 2329-9290)

EDITORIAL

Editorial: Special Issue on the Eighth Dialog System Technology Challenge	
..... S. Kim, H. Schulz, C. Gunasekara, C. Hori, A. Rastogi, and L. F. D'Haro	2434

REGULAR PAPERS

<i>Audio and Speech Modeling, Coding and Transmission</i>	
Speech Emotion Recognition Considering Nonverbal Vocalization in Affective Conversations	
..... J.-H. Hsu, M.-H. Su, C.-H. Wu, and Y.-H. Chen	1675
Multi-Source DOA Estimation in Reverberant Environments by Jointing Detection and Modeling of Time-Frequency Points	
..... M. Jia, Y. Wu, C. Bao, and C. Ritz	379
<i>Auditory Modeling and Hearing Instruments</i>	
Localization Based on Enhanced Low Frequency Interaural Level Difference	
..... M. Calis, S. van de Par, R. Heusdens, and R. C. Hendriks	3025
The Temporal Limits Encoder as a Sound Coding Strategy for Bilateral Cochlear Implants	
..... A. Kan and Q. Meng	265
<i>Acoustic Sensor Array Processing</i>	
Learning Deep Direct-Path Relative Transfer Function for Binaural Sound Source Localization	
..... B. Yang, H. Liu, and X. Li	3491
On the Design of Sparse Arrays With Frequency-Invariant Beam Pattern	
..... P. L. Son	226
Estimation Reliability Function Assisted Sound Source Localization With Enhanced Steering Vector Phase Difference	
..... L. Cheng, X. Sun, D. Yao, J. Li, and Y. Yan	421
Exploiting Temporal Context in CNN Based Multisource DOA Estimation	
..... A. Bohlender, A. Spriet, W. Tirry, and N. Madhu	1594

(Contents Continued on Page i)



Steering Study of Linear Differential Microphone Arrays	158
. <i>J. Jin, G. Huang, X. Wang, J. Chen, J. Benesty, and I. Cohen</i>	
Near-Field Superdirectivity: An Analytical Perspective	1661
. <i>D. Y. Levin, S. Markovich-Golan, and S. Gannot</i>	
Controlling Elevation and Azimuth Beamwidths With Concentric Circular Microphone Arrays	1491
. <i>R. Sharma, I. Cohen, and B. Berdugo</i>	
Multiple Source Direction of Arrival Estimations Using Relative Sound Pressure Based MUSIC	253
. <i>Y. Hu, T. D. Abhayapala, and P. N. Samarasinghe</i>	
Multiple Circular Arrays of Vector Sensors for Real-Time Sound Field Analysis	286
. <i>F. Ma, T. D. Abhayapala, and W. Zhang</i>	
Generation of Personal Sound Zones With Physical Meaningful Constraints and Conjugate Gradient Method	823
. <i>L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen</i>	
On the Design of 3D Steerable Beamformers With Uniform Concentric Circular Microphone Arrays	2764
. <i>X. Zhao, G. Huang, J. Chen, and J. Benesty</i>	
Geometry Calibration for Acoustic Transceiver Networks Based on Network Newton Distributed Optimization	1023
. <i>D. Hu, Z. Chen, and F. Yin</i>	
On the Robustness of the Superdirective Beamformer	838
. <i>X. Chen, J. Benesty, G. Huang, and J. Chen</i>	
Passive Geometry Calibration for Microphone Arrays Based on Distributed Damped Newton Optimization	118
. <i>D. Hu, Z. Chen, and F. Yin</i>	
Iterative Echo Labeling Algorithm With Convex Hull Expansion for Room Geometry Estimation	1463
. <i>S. Park and J.-W. Choi</i>	
Overview and Evaluation of Sound Event Localization and Detection in DCASE 2019	684
. <i>A. Politis, A. Mesaros, S. Adavanne, T. Heittola, and T. Virtanen</i>	
A New Class of Differential Beamformers	594
. <i>W. Yang, J. Benesty, G. Huang, and J. Chen</i>	
Minimum Processing Beamforming	2710
. <i>A. Zahedi, M. S. Pedersen, J. Østergaard, T. U. Christiansen, L. Bramsløw, and J. Jensen</i>	
A Joint Diagonalization Based Efficient Approach to Underdetermined Blind Audio Source Separation Using the Multichannel Wiener Filter	1950
. <i>N. Ito, R. Ikeshita, H. Sawada, and T. Nakatani</i>	
On Improved Training of CNN for Acoustic Source Localisation	720
. <i>E. Vargas, J. R. Hopgood, K. Brown, and K. Subr</i>	
Neural Network Adaptation and Data Augmentation for Multi-Speaker Direction-of-Arrival Estimation	1303
. <i>W. He, P. Motlicek, and J.-M. Odobez</i>	
Robustness of Acoustic Rake Filters in Minimum Variance Beamforming	3668
. <i>R. Weisman, T. Shlomo, V. Tourbabin, P. Calamia, and B. Rafaely</i>	
Multiple Acoustic Source Localization in Microphone Array Networks	334
. <i>J. Yang, X. Zhong, W. Chen, and W. Wang</i>	
Ambisonic Signal Processing DNNs Guaranteeing Rotation, Scale and Time Translation Equivariance	1449
. <i>R. Sato, K. Niwa, and K. Kobayashi</i>	
Indoor Multi-Speaker Localization Based on Bayesian Nonparametrics in the Circular Harmonic Domain	1864
. <i>K. SongGong, H. Chen, and W. Wang</i>	
Performance Analysis of the Extended Binaural MVDR Beamformer With Partial Noise Estimation	462
. <i>N. Gößling, D. Marquardt, and S. Doclo</i>	
Double-Cross-Correlation Processing for Blind Sampling-Rate and Time-Offset Estimation	1881
. <i>A. Chinaev, P. Thüne, and G.ENZNER</i>	
Beamforming with Cube Microphone Arrays Via Kronecker Product Decompositions	1774
. <i>X. Wang, J. Benesty, J. Chen, G. Huang, and I. Cohen</i>	
Differential Beamforming From the Beampattern Factorization Perspective	632
. <i>X. Zhao, J. Benesty, J. Chen, and G. Huang</i>	
Time-Frequency-Bin-Wise Linear Combination of Beamformers for Distortionless Signal Enhancement	3461
. <i>K. Yamaoka, N. Ono, and S. Makino</i>	
Sensor Imperfection Tolerance Analysis of Robust Linear Differential Microphone Arrays	2915
. <i>Z. Chen, H. Chen, and Q. Tu</i>	
<i>Audio for Multimedia</i>	
Information Fusion in Attention Networks Using Adaptive and Multi-Level Factorized Bilinear Pooling for Audio-Visual Emotion Recognition	2617
. <i>H. Zhou, J. Du, Y. Zhang, Q. Wang, Q.-F. Liu, and C.-H. Lee</i>	

Bioacoustics and Medical Acoustics

Comparison of Artificial Neural Network Types for Infant Vocalization Classification *F. Anders, M. Hlawitschka, and M. Fuchs* 54

Voice Jitter Estimation Using High-Order Synchrosqueezing Operators *J. M. Miramont, M. A. Colominas, and G. Schlotthauer* 527

Acoustic Measure of Vocal Strain Based on Glottal Airflow Periodicity *Y.-R. Chien and J. Guanason* 563

Detection and Classification of Acoustic Scenes and Events

Receptive Field Regularization Techniques for Audio Classification and Tagging With Deep Convolutional Neural Networks *K. Koutini, H. Eghbal-zadeh, and G. Widmer* 1987

Zero-Shot Audio Classification Via Semantic Embeddings *H. Xie and T. Virtanen* 1233

Temporal Dynamics of Workplace Acoustic Scenes: Egocentric Analysis and Prediction *A. Jati, A. Nadarajan, R. Peri, K. Mundnich, T. Feng, B. Girault, and S. Narayanan* 756

Guided Generative Adversarial Neural Network for Representation Learning and Audio Generation Using Fewer Labelled Audio Data *K. N. Haque, R. Rana, J. Liu, J. H. L. Hansen, N. Cummins, C. Busso, and B. W. Schuller* 2575

Privacy-Preserving Audio Classification Using Variational Information Feature Extraction *A. Nelus and R. Martin* 2864

Multi-Branch Convolutional Macaron net for Sound Event Detection *T. K. Chan and C. S. Chin* 2972

Towards Duration Robust Weakly Supervised Sound Event Detection *H. Dinkel, M. Wu, and K. Yu* 887

PSLA: Improving Audio Tagging With Pretraining, Sampling, Labeling, and Aggregation *Y. Gong, Y.-A. Chung, and J. Glass* 3292

Modeling, Analysis and Synthesis of Acoustic Environments

A Superfast Toeplitz Matrix Inversion Method for Single- and Multi-Channel Inverse Filters and Its Application to Room Equalization *M. A. Poletti and P. D. Teal* 3144

Room Acoustical Parameter Estimation From Room Impulse Responses Using Deep Neural Networks *W. Yu and W. B. Kleijn* 436

Computation of Spherical Harmonic Representations of Source Directivity Based on the Finite-Distance Signature *J. Ahrens and S. Bilbao* 83

A Room Impulse Response Measurement Method Robust Towards Nonlinearities Based on Orthogonal Periodic Sequences *A. Carini, S. Cecchi, A. Terenzi, and S. Orcioni* 3104

Robust Sound Source Tracking Using SRP-PHAT and 3D Convolutional Neural Networks *D. Diaz-Guerra, A. Miguel, and J. R. Beltran* 300

Music Information Retrieval and Music Language Processing

Hierarchical Regulated Iterative Network for Joint Task of Music Detection and Music Relative Loudness Estimation *B. Jia, J. Lv, X. Peng, Y. Chen, and S. Yang* 1

High-Resolution Piano Transcription With Pedals by Regressing Onset and Offset Times *Q. Kong, B. Li, X. Song, Y. Wan, and Y. Wang* 3707

Music Signal Analysis, Processing and Synthesis

FifthNet: Structured Compact Neural Networks for Automatic Chord Recognition *K. O'Hanlon and M. B. Sandler* 2671

Late-Reverberation Synthesis Using Interleaved Velvet-Noise Sequences *V. Välimäki and K. Prawda* 1149

Harmonic-Temporal Factor Decomposition for Unsupervised Monaural Separation of Harmonic Sounds *T. Nakamura and H. Kameoka* 68

CTC-Based Learning of Chroma Features for Score–Audio Music Retrieval *F. Zalkow and M. Müller* 2957

A Wave Digital Newton-Raphson Method for Virtual Analog Modeling of Audio Circuits with Multiple One-Port Nonlinearities *A. Bernardini, E. Bozzo, F. Fontana, and A. Sarti* 2162

Wave Digital Modeling and Implementation of Nonlinear Audio Circuits With Nullors *R. Giampiccolo, M. G. de Bari, A. Bernardini, and A. Sarti* 3267

Audio-Based Piano Performance Evaluation for Beginners With Convolutional Neural Network and Attention Mechanism *W. Wang, J. Pan, H. Yi, Z. Song, and M. Li* 1119

Active Noise Control, Echo Reduction and Feedback Reduction

Optimal Output-Constrained Active Noise Control Based on Inverse Adaptive Modeling Leak Factor Estimate *D. Shi, W.-S. Gan, B. Lam, S. Wen, and X. Shen* 1256

Robust Q -Gradient Subband Adaptive Filter for Nonlinear Active Noise Control *K.-L. Yin, Y.-F. Pu, and L. Lu* 2741

Proportionate Adaptive Filtering Algorithms Derived Using an Iterative Reweighting Framework	<i>C.-H. Lee, B. D. Rao, and H. Garudadri</i>	171
Proximal Normalized Subband Adaptive Filtering for Acoustic Echo Cancellation	<i>G. Guo, Y. Yu, R. C. de Lamare, Z. Zheng, L. Lu, and Q. Cai</i>	2174
A Novel Approach for Improved Noise Reduction Performance in Feed-Forward Active Noise Control Systems With (Loudspeaker) Saturation Non-Linearity in the Secondary Path	<i>S. Ahmed, M. Tufail, M. Rehan, T. Abbas, and A. Majid</i>	187
Affine Projection Algorithm Over Acoustic Sensor Networks for Active Noise Control	<i>M. Ferrer, M. de Diego, G. Piñero, and A. Gonzalez</i>	448
Nearest Kronecker Product Decomposition Based Linear-in-The-Parameters Nonlinear Filters	<i>S. S. Bhattacharjee and N. V. George</i>	2111
Spatial Active Noise Control in Rooms Using Higher Order Sources	<i>J. Zhang, W. Zhang, J. A. Zhang, T. D. Abhayapala, and L. Zhang</i>	3577
<i>Quality and Intelligibility Measures</i>		
Incorporating Wireless Communication Parameters Into the E-Model Algorithm	<i>D. Z. Rodríguez, D. Carrillo, M. A. Ramírez, P. H. J. Nardelli, and S. Möller</i>	956
Audibility of Group-Delay Equalization	<i>J. Liski, A. Mäkipirta, and V. Välimäki</i>	2189
Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence	<i>M. Torcoli, T. Kastner, and J. Herre</i>	1530
Speech Intelligibility Prediction Using Spectro-Temporal Modulation Analysis	<i>A. Edraki, W.-Y. Chan, J. Jensen, and D. Fogerty</i>	210
<i>Spatial Audio Recording and Reproduction</i>		
Weighted Orthogonal Vector Rejection Method for Loudspeaker-Based Binaural Audio Reproduction	<i>E. C. Hamdan and F. M. Fazi</i>	1844
A Causality-Constrained Frequency-Domain Least-Squares Filter Design Method for Crosstalk Cancellation	<i>T. Kabzinski and P. Jax</i>	2942
The Effect of Partial Time-Frequency Masking of the Direct Sound on the Perception of Reverberant Speech	<i>L. Madmoni, S. Tibor, I. Nelken, and B. Rafaely</i>	2037
Spatial Active Noise Control Based on Kernel Interpolation of Sound Field	<i>S. Koyama, J. Brunnström, H. Ito, N. Ueno, and H. Saruwatari</i>	3052
Binaural Reproduction Based on Bilateral Ambisonics and Ear-Aligned HRTFs	<i>Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely</i>	901
Sound Field Reproduction With a Cylindrical Loudspeaker Array Using First Order Wall Reflections	<i>N. Maeda, F. M. Fazi, and F.-M. Hoffmann</i>	3617
A Room Compensation Method by Modification of Reverberant Audio Objects	<i>D. Menzies, P. Coleman, and F. M. Fazi</i>	239
Fast Generation of Sound Zones Using Variable Span Trade-Off Filters in the DFT-Domain	<i>T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen</i>	363
Effects of Additive Noise in Binaural Rendering of Spherical Microphone Array Signals	<i>H. Helmholtz, D. Lou Alon, S. V. A. Garí, and J. Ahrens</i>	3642
Estimation of Spectral Notches From Pinna Meshes: Insights From a Simple Computational Model	<i>S. Spagnol, R. Miccini, M. G. Onofrei, R. Unnthorsson, and S. Serafin</i>	2683
Mixed Source Sound Field Translation for Virtual Binaural Application With Perceptual Validation	<i>L. Birnie, T. Abhayapala, V. Tourbabin, and P. Samarasinghe</i>	1188
Partially Matching Projection Decoding Method Evaluation Under Different Playback Conditions	<i>Z. Ge, L. Li, and T. Qu</i>	1411
Binaural Auralization of Microphone Array Room Impulse Responses Using Causal Wiener Filtering	<i>V. Gunnarsson and M. Sternad</i>	2899
<i>Audio Security</i>		
Non-Linear-Echo Based Anti-Collusion Mechanism for Audio Signals	<i>T. Zong, Y. Xiang, I. Natgunanathan, L. Gao, G. Hua, and W. Zhou</i>	969
Detection of Multiple Steganography Methods in Compressed Speech Based on Code Element Embedding, Bi-LSTM and CNN With Attention Mechanisms	<i>S. Li, J. Wang, P. Liu, M. Wei, and Q. Yan</i>	1556

RARS: Recognition of Audio Recording Source Based on Residual Neural Network	X. Shen, X. Shao, Q. Ge, and L. Liu	575
Desynchronization Attacks Resilient Watermarking Method Based on Frequency Singular Value Coefficient Modification	J. Zhao, T. Zong, Y. Xiang, L. Gao, W. Zhou, and G. Beliakov	2282
<i>Signal Enhancement and Restoration</i>		
Multi-Metric Optimization Using Generative Adversarial Networks for Near-End Speech Intelligibility Enhancement	H. Li and J. Yamagishi	3000
An Efficient Filter Bank Structure for Adaptive Notch Filtering and Applications	W. Wu, Y. Xiao, J. Lin, L. Ma, and K. Khorasani	3226
SNR-Based Features and Diverse Training Data for Robust DNN-Based Speech Enhancement	R. Rehr and T. Gerkmann	1937
Sparsity-Based Audio Declipping Methods: Selected Overview, New Algorithms, and Large-Scale Evaluation	C. Gaultier, S. Kitić, R. Gribonval, and N. Bertin	1174
Enhancement of Noisy Reverberant Speech Using Polynomial Matrix Eigenvalue Decomposition	V. W. Neo, C. Evers, and P. A. Naylor	3255
Towards Robust Speech Super-Resolution	H. Wang and D. L. Wang	2058
<i>Audio and Speech Source Separation</i>		
Determined BSS Based on Time-Frequency Masking and Its Application to Harmonic Vector Analysis	K. Yatabe and D. Kitamura	1609
Blind Separation for Multiple Moving Sources With Labeled Random Finite Sets	J. Ong, B. T. Vo, and S. Nordholm	2137
Reduction of Subjective Listening Effort for TV Broadcast Signals With Recurrent Neural Networks	N. L. Westhausen, R. Huber, H. Baumgartner, R. Sinha, J. Rennie, and B. T. Meyer	3541
Graph-Based Clustering of Dolphin Whistles	D. Kipnis and R. Diamant	2216
Phoneme Level Lyrics Alignment and Text-Informed Singing Voice Separation	K. Schulze-Forster, C. S. J. Doire, G. Richard, and R. Badeau	2382
Sparse Representations With Legendre Kernels for DOA Estimation and Acoustic Source Separation	M. B. Çötelı and H. Hacıhabiboğlu	2296
Evolving Multi-Resolution Pooling CNN for Monaural Singing Voice Separation	W. Yuan, B. Dong, S. Wang, M. Unoki, and W. Wang	807
Monaural Speech Separation Using Speaker Embedding From Preliminary Separation	J. Byun and J. W. Shin	2753
Group Communication With Context Codec for Lightweight Source Separation	Y. Luo, C. Han, and N. Mesgarani	1752
Time-Domain Audio Source Separation With Neural Networks Based on Multiresolution Analysis	T. Nakamura, S. Kozuka, and H. Saruwatari	1687
Conditioned Source Separation for Musical Instrument Performances	O. Slizovskaia, G. Haro, and E. Gómez	2083
<i>System Identification and Reverberation Reduction</i>		
Room Impulse Response Reshaping and Crosstalk Cancellation Using Convex Optimization	A. Mertins, M. Maass, and F. Katzberg	489
<i>Language Modelling</i>		
High-Order Pair-Wise Aspect and Opinion Terms Extraction With Edge-Enhanced Syntactic Graph Convolution	S. Wu, H. Fei, Y. Ren, B. Li, F. Li, and D. Ji	2396
Second-Order Semantic Role Labeling With Global Structural Refinement	H. Fei, S. Wu, Y. Ren, and D. Ji	1966
CSS-LM: A Contrastive Framework for Semi-Supervised Fine-Tuning of Pre-Trained Language Models	Y. Su, X. Han, Y. Lin, Z. Zhang, Z. Liu, P. Li, J. Zhou, and M. Sun	2930
Mixed Precision Low-Bit Quantization of Neural Network Language Models for Speech Recognition	J. Xu, J. Yu, S. Hu, X. Liu, and H. Meng	3679
Pre-Training With Whole Word Masking for Chinese BERT	Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang	3504
Learning Cross-Lingual Mappings in Imperfectly Isomorphic Embedding Spaces	Y. Li, K. Yu, and Y. Zhang	2630
Subword-Based Compact Reconstruction for Open-Vocabulary Neural Word Embeddings	S. Sasaki, J. Suzuki, and K. Inui	3551
<i>Multimodal Processing of Language</i>		
End-to-End Recurrent Cross-Modality Attention for Video Dialogue	Y.-W. Chu, K.-Y. Lin, C.-C. Hsu, and L.-W. Ku	2456

Efficient Learning Approach for Pronominal Anaphora and Ellipsis Identification and Resolution in Arabic Texts	3335
..... <i>S. Mathlouthi Bouzid and C. Ben Othmane Zribi</i>	
Synthesizing Spoken Descriptions of Images	3242
..... <i>X. Wang, J. van der Hout, J. Zhu, M. Hasegawa-Johnson, and O. Scharenborg</i>	
<i>Machine Translation for Spoken and Written Language</i>	
Neural Machine Translation With Explicit Phrase Alignment	1001
..... <i>J. Zhang, H. Luan, M. Sun, F. Zhai, J. Xu, and Y. Liu</i>	
Detecting Source Contextual Barriers for Understanding Neural Machine Translation	1001
..... <i>G. Li, L. Liu, C. Zhu, R. Wang, T. Zhao, and S. Shi</i>	
Preordering Encoding on Transformer for Translation	644
..... <i>Y. Kawara, C. Chu, and Y. Arase</i>	
Modeling Future Cost for Neural Machine Translation	770
..... <i>C. Duan, K. Chen, R. Wang, M. Utiyama, E. Sumita, C. Zhu, and T. Zhao</i>	
Exploiting Translation Model for Parallel Corpus Mining	2829
..... <i>C. Leong, X. Liu, D. F. Wong, and L. S. Chao</i>	
Improving Automatic Speech Recognition and Speech Translation via Word Embedding Prediction	93
..... <i>S.-P. Chuang, A. H. Liu, T.-W. Sung, and H.-yi Lee</i>	
Attending From Foresight: A Novel Attention Mechanism for Neural Machine Translation	2606
..... <i>X. Li, L. Liu, Z. Tu, G. Li, S. Shi, and M. Q.-H. Meng</i>	
Adaptive Adapters: An Efficient Way to Incorporate BERT Into Neural Machine Translation	1740
..... <i>J. Guo, Z. Zhang, L. Xu, B. Chen, and E. Chen</i>	
<i>Language Understanding and Computational Semantics</i>	
Flexibly Focusing on Supporting Facts, Using Bridge Links, and Jointly Training Specialized Modules for Multi-Hop Question Answering	3216
..... <i>T. Alkhalidi, C. Chu, and S. Kurohashi</i>	
A Multiple-Integration Encoder for Multi-Turn Text-to-SQL Semantic Parsing	1503
..... <i>R.-Z. Wang, Z.-H. Ling, J.-B. Zhou, and Y. Hu</i>	
Sarcasm Detection with Commonsense Knowledge	3192
..... <i>J. Li, H. Pan, Z. Lin, P. Fu, and W. Wang</i>	
Generating Images From Spoken Descriptions	850
..... <i>X. Wang, T. Qiao, J. Zhu, A. Hanjalic, and O. Scharenborg</i>	
Addressing Extraction and Generation Separately: Keyphrase Prediction With Pre-Trained Language Models	3180
..... <i>R. Liu, Z. Lin, and W. Wang</i>	
Adaptive Convolution for Semantic Role Labeling	782
..... <i>K. Munir, H. Zhao, and Z. Li</i>	
A Knowledge Graph Embedding Approach for Metaphor Processing	406
..... <i>W. Song, J. Guo, R. Fu, T. Liu, and L. Liu</i>	
Randomly Wired Network Based on RoBERTa and Dialog History Attention for Response Selection	2437
..... <i>B. Kim, J. Seo, and M.-W. Koo</i>	
Hierarchical Neighbor Propagation With Bidirectional Graph Attention Network for Relation Prediction	1762
..... <i>Z. Xie, R. Zhu, J. Liu, G. Zhou, and J. X. Huang</i>	
Semantic Change Detection With Gaussian Word Embeddings	3349
..... <i>A. Yüksel, B. Uğurlu, and A. Koç</i>	
HDRS: Hindi Dialogue Restaurant Search Corpus for Dialogue State Tracking in Task-Oriented Environment	2517
..... <i>S. Malviya, R. Mishra, S. K. Barnwal, and U. S. Tiwary</i>	
Contrastive Information Extraction With Generative Transformer	3077
..... <i>N. Zhang, H. Ye, S. Deng, C. Tan, M. Chen, S. Huang, F. Huang, and H. Chen</i>	
Learning Fine-Grained Fact-Article Correspondence in Legal Cases	3694
..... <i>J. Ge, Y. Huang, X. Shen, C. Li, and W. Hu</i>	
Efficient Estimate of Sentence's Representation Based on the Difference Semantics Model	3384
..... <i>X. Liao, Y. Huang, Y. Wei, C. Zhang, F. Wang, and Y. Wang</i>	
Review and Arrange: Curriculum Learning for Natural Language Understanding	3307
..... <i>L. Zhang, Z. Mao, B. Xu, Q. Wang, and Y. Zhang</i>	
Analyzing Multimodal Sentiment Via Acoustic- and Visual-LSTM With Channel-Aware Temporal Convolution Network	1424
..... <i>S. Mai, S. Xing, and H. Hu</i>	
A Graph-to-Sequence Learning Framework for Summarizing Opinionated Texts	1650
..... <i>P. Wei, J. Zhao, and W. Mao</i>	
FSPRM: A Feature Subsequence Based Probability Representation Model for Chinese Word Embedding	1702
..... <i>Y. Zhang, Y. Liu, J. Zhu, and X. Wu</i>	
Conversational Semantic Role Labeling	2465
..... <i>K. Xu, H. Wu, L. Song, H. Zhang, L. Song, and D. Yu</i>	
Improving Skip-Gram Embeddings Using BERT	1318
..... <i>Y. Wang, L. Cui, and Y. Zhang</i>	
<i>Discourse and Dialog</i>	
Domain Adaptive Meta-Learning for Dialogue State Tracking	2493
..... <i>J. Zeng, Y. Yin, Y. Liu, Y. Ge, and J. Su</i>	
Multi-Turn Dialogue Reading Comprehension With Pivot Turns and Knowledge	1161
..... <i>Z. Zhang, J. Li, and H. Zhao</i>	

Bridging Text and Video: A Universal Multimodal Transformer for Audio-Visual Scene-Aware Dialog	2476
. Z. Li, Z. Li, J. Zhang, Y. Feng, and J. Zhou	
Knowing Where to Leverage: Context-Aware Graph Convolutional Network With an Adaptive Fusion Layer for Contextual Spoken Language Understanding	1280
. L. Qin, W. Che, M. Ni, Y. Li, and T. Liu	
Domain-Aware Dialogue State Tracker for Multi-Domain Dialogue Systems	866
. V. Balaraman and B. Magnini	
Medical Term and Status Generation From Chinese Clinical Dialogue With Multi-Granularity Transformer	3362
. M. Li, L. Xiang, X. Kang, Y. Zhao, Y. Zhou, and C. Zong	
Learning to Generate Explainable Plots for Neural Story Generation	585
. G. Chen, Y. Liu, H. Luan, M. Zhang, Q. Liu, and M. Sun	
D-Score: Holistic Dialogue Evaluation Without Reference	2502
. C. Zhang, G. Lee, L. F. D’Haro, and H. Li	
Label and Context Augmentation for Response Selection at DSTC8	2541
. M. Jeong, S. Choi, J. Yeo, and S.-won Hwang	
Corruption Is Not All Bad: Incorporating Discourse Structure INTO Pre-Training via Corruption for Essay Scoring	2202
. F. S. Mim, N. Inoue, P. Reisert, H. Ouchi, and K. Inui	
Deep Contextualized Utterance Representations for Response Selection and Dialogue Analysis	2443
. J.-C. Gu, T. Li, Z.-H. Ling, Q. Liu, Z. Su, Y.-P. Ruan, and X. Zhu	
Learning Context-Aware Convolutional Filters for Implicit Discourse Relation Classification	2421
. K. Munir, H. Zhao, and Z. Li	
GRT: Generative-Retrieval Transformers for Data-Efficient Dialogue Domain Adaptation	2484
. I. Shalymov, A. Sordoni, A. Atkinson, and H. Schulz	
Predict-Then-Decide: A Predictive Approach for Wait or Answer Task in Dialogue Systems	3012
. Z. Lin, S. Cui, G. Li, X. Kang, F. Ji, F. Li, Z. Zhao, H. Chen, and Y. Zhang	
Generating Responses With a Given Syntactic Pattern in Chinese Dialogues	2888
. Y. Zhou, X. Zheng, and X. Huang	
PROTOTYPE-TO-STYLE: Dialogue Generation With Style-Aware Editing on Retrieval Memory	2152
. Y. Su, Y. Wang, D. Cai, S. Baker, A. Korhonen, and N. Collier	
Overview of the Eighth Dialog System Technology Challenge: DSTC8	2529
. S. Kim, M. Galley, C. Gunasekara, S. Lee, A. Atkinson, B. Peng, H. Schulz, J. Gao, J. Li, M. Adada, M. Huang, L. Lastras, J. K. Kummerfeld, W. S. Lasecki, C. Hori, A. Cherian, T. K. Marks, A. Rastogi, X. Zang, S. Sunkara, and R. Gupta	
<i>Spoken Document Retrieval and Text Mining</i>	
Regularized Phrase-Based Topic Model for Automatic Question Classification With Domain-Agnostic Class Labels	3604
. S. Supraja, A. W. H. Khong, and S. Tatinati	
Multi-Task Sequence Tagging for Emotion-Cause Pair Extraction Via Tag Distribution Refinement	2339
. C. Fan, C. Yuan, L. Gui, Y. Zhang, and R. Xu	
Knowledge Enhanced Fact Checking and Verification	3132
. B. Zhu, X. Zhang, M. Gu, and Y. Deng	
A Unified Target-Oriented Sequence-to-Sequence Model for Emotion-Cause Pair Extraction	2779
. Z. Cheng, Z. Jiang, Y. Yin, N. Li, and Q. Gu	
Deep Selective Memory Network With Selective Attention and Inter-Aspect Modeling for Aspect Level Sentiment Classification	1093
. P. Lin, M. Yang, and J. Lai	
Investigating Typed Syntactic Dependencies for Targeted Sentiment Classification Using Graph Attention Neural Network	503
. X. Bai, P. Liu, and Y. Zhang	
Deformable Self-Attention for Text Classification	1570
. Q. Ma, J. Yan, Z. Lin, L. Yu, and Z. Chen	
History Reuse and Bag-of-Words Loss for Long Summary Generation	2551
. Q. Liu, L. Chen, Y. Yuan, and H. Wu	
Sentiment Time Series Calibration for Event Detection	2407
. J. Wu, L. Shang, and X. Gao	
LSBert: Lexical Simplification Based on BERT	3064
. J. Qiang, Y. Li, Y. Zhu, Y. Yuan, Y. Shi, and X. Wu	
Relation Extraction in Dialogues: A Deep Learning Model Based on the Generality and Specialty of Dialogue Text	2015
. M. Zhou, D. Ji, and F. Li	
Audio-Aware Spoken Multiple-Choice Question Answering With Pre-Trained Language Models	3170
. C.-C. Kuo, K.-Y. Chen, and S.-B. Luo	
<i>Segmentation, Tagging and Parsing</i>	
A Multi-Agent Communication Based Model for Nested Named Entity Recognition	2123
. C. Li, G. Wang, J. Cao, and Y. Cai	
Deep Graph-Based Character-Level Chinese Dependency Parsing	1329
. L. Wu and M. Zhang	
Corpus-Aware Graph Aggregation Network for Sequence Labeling	2048
. H. Chen, Q. Ma, L. Yu, Z. Lin, and J. Yan	
A Joint Model for Named Entity Recognition With Sentence-Level Entity Type Attentions	1438
. T. Qian, M. Zhang, Y. Lou, and D. Hua	

Language Acquisition and Learning

Domain-Shift Conditioning Using Adaptable Filtering Via Hierarchical Embeddings for Robust Chinese Spell Check *M. Nguyen, G. H. Ngo, and N. F. Chen* 2027

Chinese Lexical Simplification *J. Qiang, X. Lu, Y. Li, Y. Yuan, and X. Wu* 1819

Speech Production

F0 Perturbation Due to Articulatory Movements: Filtering, Characterization and Applications *H. M. Torres, M. Güemes, J. A. Gurlekian, and D. A. Evin* 1977

Synthesis and Analysis-By-Synthesis of Modulated Diplophonic Glottal Area Waveforms *P. Aichinger and F. Pernkopf* 914

Lyric or Dramatic - Vibrato Analysis for Voice Type Classification in Professional Opera Singers *M. Müller, T. Schulz, T. Ermakova, and P. P. Caffier* 943

Speech Perception and Psychoacoustics

Directly Comparing the Listening Strategies of Humans and Machines *V. A. Trinh and M. Mandel* 312

Multi-Tone Phase Coding of Interaural Time Difference for Sound Source Localization With Spiking Neural Networks *Z. Pan, M. Zhang, J. Wu, J. Wang, and H. Li* 2656

Automatic Detection of Affective Flattening in Schizophrenia: Acoustic Correlates to Sound Waves and Auditory Perception *F. He, L. He, J. Zhang, Y. Li, and X. Xiong* 3321

Listening Enhancement in Noisy Environments: Solutions in Time and Frequency Domain *M. Niermann and P. Vary* 699

Speech Analysis

Robust Voice Feature Selection Using Interval Type-2 Fuzzy AHP for Automated Diagnosis of Parkinson’s Disease *H. Azadi, M.-R. Akbarzadeh-T, H.-R. Kobravi, and A. Shoeibi* 2792

Overlapping Speaker Segmentation Using Multiple Hypothesis Tracking of Fundamental Frequency *A. O. T. Hogg, C. Evers, A. H. Moore, and P. A. Naylor* 1479

Block-Based High Performance CNN Architectures for Frame-Level Overlapping Speech Detection *M. Yousefi and J. H. L. Hansen* 28

Speech Emotion Recognition Using Sequential Capsule Networks *X. Wu, Y. Cao, H. Lu, S. Liu, D. Wang, Z. Wu, X. Liu, and H. Meng* 3280

The Detection of Parkinson’s Disease From Speech Using Voice Source Information *N. P. Narendra, B. Schuller, and P. Alku* 1925

Recurrent Neural Networks and Acoustic Features for Frame-Level Signal-to-Noise Ratio Estimation *H. Li, D. L. Wang, X. Zhang, and G. Gao* 2878

Novel Architectures for Unsupervised Information Bottleneck Based Speaker Diarization of Meetings *N. Dawalatabad, S. Madikeri, C. C. Sekhar, and H. A. Murthy* 14

CTNet: Conversational Transformer Network for Emotion Recognition *Z. Lian, B. Liu, and J. Tao* 985

Hybrid Speech and Text Analysis Methods for Speaker Change Detection *O. H. Anidjar, I. Lapidot, C. Hajaj, A. Dvir, and I. Gilad* 2324

F_0 -Noise-Robust Glottal Source and Vocal Tract Analysis Based on ARX-LF Model *Y. Li, J. Tao, D. Erickson, B. Liu, and M. Akagi* 3375

Ensemble Bag-of-Audio-Words Representation Improves Paralinguistic Classification Accuracy *G. Gosztolya and R. Busa-Fekete* 477

FluentNet: End-to-End Detection of Stuttered Speech Disfluencies With Deep Learning *T. Kourkounakis, A. Hajavi, and A. Etemad* 2986

Language-Independent Approach for Automatic Computation of Vowel Articulation Features in Dysarthric Speech Assessment *Y. Liu, N. Penttilä, T. Ihalainen, J. Lintula, R. Convey, and O. Räsänen* 2228

Speech Synthesis and Generation

UnitNet: A Sequence-to-Sequence Acoustic Model for Concatenative Speech Synthesis *X. Zhou, Z.-H. Ling, and L.-R. Dai* 2643

Perceptual-Similarity-Aware Deep Speaker Representation Learning for Multi-Speaker Generative Modeling *Y. Saito, S. Takamichi, and H. Saruwatari* 1033

Any-to-Many Voice Conversion With Location-Relative Sequence-to-Sequence Modeling *S. Liu, Y. Cao, D. Wang, X. Wu, X. Liu, and H. Meng* 1717

Native-Nonnative Voice Conversion by Residual Warping in a Sparse, Anchor-Based Representation	<i>C. Liberatoro</i>	3040
Language Agnostic Speaker Embedding for Cross-Lingual Personalized Speech Generation	<i>Y. Zhou, X. Tian, and H. Li</i>	3427
An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning	<i>B. Sisman, J. Yamagishi, S. King, and H. Li</i>	132
Robustness of Speech Spoofing Detectors Against Adversarial Post-Processing of Voice Conversion	<i>Y.-Y. Ding, H.-J. Lin, L.-J. Liu, Z.-H. Ling, and Y. Hu</i>	3415
Expressive TTS Training With Frame and Style Reconstruction Loss	<i>R. Liu, B. Sisman, G. Gao, and H. Li</i>	1806
Exploiting Morphological and Phonological Features to Improve Prosodic Phrasing for Mongolian Speech Synthesis	<i>R. Liu, B. Sisman, F. Bao, J. Yang, G. Gao, and H. Li</i>	274
Sinsy: A Deep Neural Network-Based Singing Voice Synthesis System	<i>Y. Hono, K. Hashimoto, K. Oura, Y. Nankaku, and K. Tokuda</i>	2803
Gamma Boltzmann Machine for Audio Modeling	<i>T. Nakashika and K. Yatabe</i>	2591
Pretraining Techniques for Sequence-to-Sequence Voice Conversion	<i>W.-C. Huang, T. Hayashi, Y.-C. Wu, H. Kameoka, and T. Toda</i>	745
Extracting and Predicting Word-Level Style Variations for Speech Synthesis	<i>Y.-J. Zhang and Z.-H. Ling</i>	1582
Exemplar-Based Emotive Speech Synthesis	<i>X. Wu, Y. Cao, H. Lu, S. Liu, S. Kang, Z. Wu, X. Liu, and H. Meng</i>	874
Quasi-Periodic WaveNet: An Autoregressive Raw Waveform Generative Model With Pitch-Dependent Dilated Convolution Neural Network	<i>Y.-C. Wu, T. Hayashi, P. L. Tobing, K. Kobayashi, and T. Toda</i>	1134
Many-to-Many Voice Transformer Network	<i>H. Kameoka, W.-C. Huang, K. Tanaka, T. Kaneko, N. Hojo, and T. Toda</i>	656
Converting Foreign Accent Speech Without a Reference	<i>G. Zhao, S. Ding, and R. Gutierrez-Osuna</i>	2367
Quasi-Periodic Parallel WaveGAN: A Non-Autoregressive Raw Waveform Generative Model With Pitch-Dependent Dilated Convolution Neural Network	<i>Y.-C. Wu, T. Hayashi, T. Okamoto, H. Kawai, and T. Toda</i>	792
Transfer Learning From Speech Synthesis to Voice Conversion With Non-Parallel Training Data	<i>M. Zhang, Y. Zhou, L. Zhao, and H. Li</i>	1290
Speech Reconstruction With Reminiscent Sound Via Visual Voice Memory	<i>J. Hong, M. Kim, S. J. Park, and Y. M. Ro</i>	3654
<i>Speech Enhancement and Separation</i>		
Speech Enhancement via Attention Masking Network (SEAMNET): An End-to-End System for Joint Suppression of Noise and Reverberation	<i>B. J. Borgström and M. S. Brandstein</i>	515
Multichannel Iterative Noise Reduction Filters in the Short-Time-Fourier-Transform Domain Based on Kronecker Product Decomposition	<i>X. Wang, J. Chen, X. Chen, J. Guo, and Q. Xiang</i>	2725
DNN-Based Mask Estimation for Distributed Speech Enhancement in Spatially Unconstrained Microphone Arrays	<i>N. Furnon, R. Serizel, S. Essid, and I. Illina</i>	2310
TAU-Net: Temporal Activation U-Net Shared With Nonnegative Matrix Factorization for Speech Enhancement in Unseen Noise Environments	<i>K. M. Jeon, G. W. Lee, N. K. Kim, and H. K. Kim</i>	3400
A Deep Adaptation Network for Speech Enhancement: Combining a Relativistic Discriminator With Multi-Kernel Maximum Mean Discrepancy	<i>J. Cheng, R. Liang, Z. Liang, L. Zhao, C. Huang, and B. Schuller</i>	41
Sensor Selection for Relative Acoustic Transfer Function Steered Linearly-Constrained Beamformers	<i>J. Zhang, J. Du, and L.-R. Dai</i>	1220
A Cross-Entropy-Guided Measure (CEGM) for Assessing Speech Recognition Performance and Optimizing DNN-Based Speech Enhancement	<i>L. Chai, J. Du, Q.-F. Liu, and C.-H. Lee</i>	106
A Study on Reference Microphone Selection for Multi-Microphone Speech Enhancement	<i>J. Zhang, H. Chen, L.-R. Dai, and R. C. Hendriks</i>	671
Impulsive Noise Detection for Speech Enhancement in HHT Domain	<i>C. Medina, R. Coelho, and L. Zão</i>	2244
Speech Enhancement Based on Modulation-Domain Parametric Multichannel Kalman Filtering	<i>W. Xue, A. H. Moore, M. Brookes, and P. A. Naylor</i>	393
Convolutional Prediction for Monaural Speech Dereverberation and Noisy-Reverberant Speaker Separation	<i>Z.-Q. Wang, G. Wichern, and J. L. Roux</i>	3476
Minimum-Volume Multichannel Nonnegative Matrix Factorization for Blind Audio Source Separation	<i>J. Wang, S. Guan, S. Liu, and X.-L. Zhang</i>	3089
An Optimal Envelope-Based Noise Reduction Method for Cochlear Implants: An Upper Bound Performance Investigation	<i>R. A. Chiea, M. H. Costa, and J. A. Cordioli</i>	1729
Quantization-Aware Binaural MWF Based Noise Reduction Incorporating External Wireless Devices	<i>J. Zhang and C. Li</i>	3118

Nonlinear Spatial Filtering in Multichannel Speech Enhancement	<i>K. Tesch and T. Gerkmann</i>	1795
Multi-Channel Multi-Frame ADL-MVDR for Target Speech Separation.		
. <i>Z. Zhang, Y. Xu, M. Yu, S.-X. Zhang, L. Chen, D. S. Williamson, and D. Yu</i>		3526
Robust Constrained MFMVDR Filters for Single-Channel Speech Enhancement Based on Spherical Uncertainty Set	<i>D. Fischer and S. Doclo</i>	618
Multi-microphone Complex Spectral Mapping for Utterance-wise and Continuous Speech Separation.		
. <i>Z.-Q. Wang, P. Wang, and D. L. Wang</i>		2001
Multichannel Blind Source Separation Based on Evanescent-Region-Aware Non-Negative Tensor Factorization in Spherical Harmonic Domain	<i>Y. Mitsufoji, N. Takamune, S. Koyama, and H. Saruwatari</i>	607
On the Design of Differential Kronecker Product Beamformers	<i>G. Itzhak, J. Benesty, and I. Cohen</i>	1397
Towards Model Compression for Deep Learning Based Speech Enhancement	<i>K. Tan and D. L. Wang</i>	1785
PhaseDCN: A Phase-Enhanced Dual-Path Dilated Convolutional Network for Single-Channel Speech Enhancement		
. <i>L. Zhang, M. Wang, Q. Zhang, X. Wang, and M. Liu</i>		2561
Wavesplit: End-to-End Speech Separation by Speaker Clustering	<i>N. Zeghidour and D. Grangier</i>	2840
Deep Learning Based Real-Time Speech Enhancement for Dual-Microphone Mobile Phones.		
. <i>K. Tan, X. Zhang, and D. L. Wang</i>		1853
An Overview of Deep-Learning-Based Audio-Visual Speech Enhancement and Separation		
. <i>D. Michelsanti, Z.-H. Tan, S.-X. Zhang, Y. Xu, M. Yu, D. Yu, and J. Jensen</i>		1368
Dense CNN With Self-Attention for Time-Domain Speech Enhancement	<i>A. Pandey and D. L. Wang</i>	1270
Speech Enhancement Using Multi-Stage Self-Attentive Temporal Convolutional Networks		
. <i>J. Lin, A. J. de L. van Wijngaarden, K.-C. Wang, and M. C. Smith</i>		3440
Speaker Separation Using Speaker Inventories and Estimated Speech		
. <i>P. Wang, Z. Chen, D. L. Wang, J. Li, and Y. Gong</i>		537
Two Heads are Better Than One: A Two-Stage Complex Spectral Mapping Approach for Monaural Speech Enhancement	<i>A. Li, W. Liu, C. Zheng, C. Fan, and X. Li</i>	1829
<i>Acoustic Modeling for Automatic Speech Recognition</i>		
Learning Waveform-Based Acoustic Models Using Deep Variational Convolutional Neural Networks		
. <i>D. Oglic, Z. Cvetkovic, and P. Sollich</i>		2850
Asynchronous Decentralized Distributed Training of Acoustic Models		
. <i>X. Cui, W. Zhang, A. Kayi, M. Liu, U. Finkler, B. Kingsbury, G. Saon, and D. Kung</i>		3565
HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units		
. <i>W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhota, R. Salakhutdinov, and A. Mohamed</i>		3451
Bayesian Learning of LF-MMI Trained Time Delay Neural Networks for Speech Recognition		
. <i>S. Hu, X. Xie, S. Liu, J. Yu, Z. Ye, M. Geng, X. Liu, and H. Meng</i>		1514
Counterfactually Fair Automatic Speech Recognition	<i>L. Sari, M. Hasegawa-Johnson, and C. D. Yoo</i>	3515
TERA: Self-Supervised Learning of Transformer Encoder Representation for Speech		
. <i>A. T. Liu, S.-W. Li, and H.-yi Lee</i>		2351
<i>Robust Speech Recognition</i>		
Gated Recurrent Fusion With Joint Training Framework for Robust End-to-End Speech Recognition		
. <i>C. Fan, J. Yi, J. Tao, Z. Tian, B. Liu, and Z. Wen</i>		198
Multimodal Emotion Recognition With Temporal and Semantic Consistency		
. <i>B. Chen, Q. Cao, M. Hou, Z. Zhang, G. Lu, and D. Zhang</i>		3592
A Novel Loss Function and Training Strategy for Noise-Robust Keyword Spotting		
. <i>I. López-Espejo, Z.-H. Tan, and J. Jensen</i>		2254
Audio-Visual Multi-Channel Integration and Recognition of Overlapped Speech		
. <i>J. Yu, S.-X. Zhang, B. Wu, S. Liu, S. Hu, M. Geng, X. Liu, H. Meng, and D. Yu</i>		2067
Convolutional Maximum-Likelihood Distortionless Response Beamforming With Steering Vector Estimation for Robust Speech Recognition	<i>B. J. Cho and H.-M. Park</i>	1352
<i>Speech Adaptation/Normalization</i>		
Auxiliary Networks for Joint Speaker Adaptation and Speaker Change Detection		
. <i>L. Sari, M. Hasegawa-Johnson, and S. Thomas</i>		324
Bayesian Learning for Deep Neural Network Adaptation	<i>X. Xie, X. Liu, T. Lee, and L. Wang</i>	2096

Vocal Tract Length Estimation Using Accumulated Means of Formants and Its Effects on Speaker-Normalization	1049
..... T. Sakata, N. Ikeda, Y. Ueda, and A. Watanabe	
<i>General Topics in Speech Recognition</i>	
Tackling Perception Bias in Unsupervised Phoneme Discovery Using DPGMM-RNN Hybrid Model and Functional Load	348
..... B. Wu, S. Sakti, J. Zhang, and S. Nakamura	
Multi-Granularity Sequence Alignment Mapping for Encoder-Decoder Based End-to-End ASR	2816
..... J. Tang, J. Zhang, Y. Song, I. McLoughlin, and L.-R. Dai	
Voice Activity Detection in the Wild: A Data-Driven Approach Using Teacher-Student Training	1542
..... H. Dinkel, S. Wang, X. Xu, M. Wu, and K. Yu	
Fast End-to-End Speech Recognition Via Non-Autoregressive Models and Cross-Modal Knowledge Transferring From BERT	1897
..... Y. Bai, J. Yi, J. Tao, Z. Tian, Z. Wen, and S. Zhang	
Improved Acoustic Word Embeddings for Zero-Resource Languages Using Multilingual Transfer	1107
..... H. Kamper, Y. Matusevych, and S. Goldwater	
Gated Recurrent Context: Softmax-Free Attention for Online Encoder-Decoder Speech Recognition	710
..... H. Lee, W. H. Kang, S. J. Cheon, H. Kim, and N. S. Kim	
Recent Progress in the CUHK Dysarthric Speech Recognition System	2267
..... S. Liu, M. Geng, S. Hu, X. Xie, M. Cui, J. Yu, X. Liu, and H. Meng	
Cognitive Load Estimation From Speech Commands to Simulated Aircraft	1011
..... M. Vukovic, M. Stolar, and M. Lech	
Multimodal Representations for Synchronized Speech and Real-Time MRI Video Processing	1912
..... Ö.ykii. D. Köse and M. Saraçlar	
TutorNet: Towards Flexible Knowledge Distillation for End-to-End Speech Recognition	1626
..... J. W. Yoon, H. Lee, H. Y. Kim, W. I. Cho, and N. S. Kim	
Integrating Knowledge Into End-to-End Speech Recognition From External Text-Only Data	1340
..... Y. Bai, J. Yi, J. Tao, Z. Wen, Z. Tian, and S. Zhang	
<i>Large Vocabulary Continuous Recognition/Search</i>	
Keyword Search Using Attention-Based End-to-End ASR and Frame-Synchronous Phoneme Alignments	3202
..... R. Yang, G. Cheng, H. Miao, T. Li, P. Zhang, and Y. Yan	
<i>Speaker Recognition and Characterization</i>	
Target Speaker Verification With Selective Auditory Attention for Single and Multi-Talker Speech	2696
..... C. Xu, W. Rao, J. Wu, and H. Li	
Phoneme-Unit-Specific Time-Delay Neural Network for Speaker Verification	1243
..... X. Chen and C. Bao	
On the Distribution of Speaker Verification Scores: Generative Models for Unsupervised Calibration	547
..... S. Cumani	
Meta-Learning With Latent Space Clustering in Generative Adversarial Network for Speaker Diarization	1204
..... M. Pal, M. Kumar, R. Peri, T. J. Park, S. H. Kim, C. Lord, S. Bishop, and S. Narayanan	
Coupling a Generative Model With a Discriminative Learning Framework for Speaker Verification	3631
..... X. Lu, P. Shen, Y. Tsao, and H. Kawai	
Analysis and Calibration of Lombard Effect and Whisper for Speaker Recognition	927
..... F. Kelly and J. H.L. Hansen	
Audio-Visual Deep Neural Network for Robust Person Verification	1079
..... Y. Qian, Z. Chen, and S. Wang	
Self-Supervised Representation Learning With Path Integral Clustering for Speaker Diarization	1639
..... P. Singh and S. Ganapathy	
Modified Magnitude-Phase Spectrum Information for Spoofing Detection	1065
..... J. Yang, H. Wang, R. K. Das, and Y. Qian	
Deep Normalization for Speaker Vectors	733
..... Y. Cai, L. Li, A. Abel, X. Zhu, and D. Wang	
