IEEE/ACM Transactions on Audio, Speech, and Language Processing

4.1
Impact Factor

0.00915
Eigenfactor

0.96
Article Influence Score

11.3
CiteScore
Powered by Scopus

Title History

Popular Documents - April 2025

Popular Article Feed

Includes the 50 most frequently accessed documents for this publication.

Export | Email Selected Results

Showing **1-50** of **50**

Date
April 2025

1. **End-to-End Speech Recognition: A Survey**
Rohit Prabhavalkar; Takaaki Hori; Tara N. Sainath; Ralf Schlüter; Shinji Watanabe
Publication Year: 2024 , Page(s): 325 - 351
Cited by: Papers (67)
Abstract HTML

2. **Conv-TasNet: Surpassing Ideal Time–Frequency Magnitude Masking for Speech Separation**
Yi Luo; Nima Mesgarani
Publication Year: 2019 , Page(s): 1256 - 1266
Cited by: Papers (1229) | Patents (1)
Abstract HTML

3. **Speech Enhancement and Dereverberation With Diffusion-Based Generative Models**
Julius Richter; Simon Welker; Jean-Marie Lemercier; Bunlong Lay; Timo Gerkmann
Publication Year: 2023 , Page(s): 2351 - 2364
Cited by: Papers (104)
Abstract HTML

4. **An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning**
Berrak Sisman; Junichi Yamagishi; Simon King; Haizhou Li
Publication Year: 2021 , Page(s): 132 - 157
Cited by: Papers (199)
Abstract HTML

5. **HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units**
Wei-Ning Hsu; Benjamin Bolte; Yao-Hung Hubert Tsai; Kushal Lakhotia; Ruslan Salakhutdinov; Abdelrahman Mohamed
Publication Year: 2021 , Page(s): 3451 - 3460
Cited by: Papers (1177)
Abstract HTML

6. **SoundStream: An End-to-End Neural Audio Codec**
Neil Zeghidour; Alejandro Luebs; Ahmed Omran; Jan Skoglund; Marco Tagliasacchi
Publication Year: 2022 , Page(s): 495 - 507
Cited by: Papers (248)

Abstract HTML

Music ControlNet: Multiple Time-Varying Controls for Music Generation
Shih-Lun Wu; Chris Donahue; Shinji Watanabe; Nicholas J. Bryan
Publication Year: 2024 , Page(s): 2692 - 2703
Cited by: Papers (10)

Abstract HTML

A Time-Frequency Attention Module for Neural Speech Enhancement
Qiquan Zhang; Xinyuan Qian; Zhaoheng Ni; Aaron Nicolson; Eliathamby Ambikairajah; Haizhou Li
Publication Year: 2023 , Page(s): 462 - 475
Cited by: Papers (27)

Abstract HTML

ZMM-TTS: Zero-Shot Multilingual and Multispeaker Speech Synthesis Conditioned on Self-Supervised Discrete Speech Representations
Cheng Gong; Xin Wang; Erica Cooper; Dan Wells; Longbiao Wang; Jianwu Dang; Korin Richmond; Junichi Yamagishi
Publication Year: 2024 , Page(s): 4036 - 4051
Cited by: Papers (7)

Abstract HTML

Pre-Training With Whole Word Masking for Chinese BERT
Yiming Cui; Wanxiang Che; Ting Liu; Bing Qin; Ziqing Yang
Publication Year: 2021 , Page(s): 3504 - 3514
Cited by: Papers (635)

Abstract HTML

Self-Supervised ASR Models and Features for Dysarthric and Elderly Speech Recognition
Shujie Hu; Xurong Xie; Mengzhe Geng; Zengrui Jin; Jiajun Deng; Guinan Li; Yi Wang; Mingyu Cui; Tianzi Wang; Helen Meng; Xunying Liu
Publication Year: 2024 , Page(s): 3561 - 3575
Cited by: Papers (3)

Abstract HTML

CMGAN: Conformer-Based Metric-GAN for Monaural Speech Enhancement
Sherif Abdulatif; Ruizhe Cao; Bin Yang
Publication Year: 2024 , Page(s): 2477 - 2493
Cited by: Papers (25)

Abstract HTML

PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition
Qiuqiang Kong; Yin Cao; Turab Iqbal; Yuxuan Wang; Wenwu Wang; Mark D. Plumbley
Publication Year: 2020 , Page(s): 2880 - 2894
Cited by: Papers (616)

Abstract HTML

Masked Modeling Duo: Towards a Universal Audio Pre-Training Framework
Daisuke Niizumi; Daiki Takeuchi; Yasunori Ohishi; Noboru Harada; Kunio Kashino
Publication Year: 2024 , Page(s): 2391 - 2406
Cited by: Papers (6)

Abstract HTML

AudioLDM 2: Learning Holistic Audio Generation With Self-Supervised Pretraining
Haohe Liu; Yi Yuan; Xubo Liu; Xinhao Mei; Qiuqiang Kong; Qiao Tian; Yuping Wang; Wenwu Wang; Yuxuan Wang; Mark D. Plumbley
Publication Year: 2024 , Page(s): 2871 - 2883

Cited by: Papers (53)

Abstract HTML

SpatialNet: Extensively Learning Spatial Information for Multichannel Joint Speech Separation, Denoising and Dereverberation
Changsheng Quan; Xiaofei Li
Publication Year: 2024 , Page(s): 1310 - 1323
Cited by: Papers (23)

Abstract HTML

Disentangling Prosody Representations With Unsupervised Speech Reconstruction
Leyuan Qu; Taihao Li; Cornelius Weber; Theresa Pekarek-Rosin; Fuji Ren; Stefan Wermter
Publication Year: 2024 , Page(s): 39 - 54
Cited by: Papers (2)

Abstract HTML

Masked Graph Learning With Recurrent Alignment for Multimodal Emotion Recognition in Conversation
Tao Meng; Fuchen Zhang; Yuntao Shou; Hongen Shao; Wei Ai; Keqin Li
Publication Year: 2024 , Page(s): 4298 - 4312
Cited by: Papers (5)

Abstract HTML

Dynamic Convolutional Neural Networks as Efficient Pre-Trained Audio Models
Florian Schmid; Khaled Koutini; Gerhard Widmer
Publication Year: 2024 , Page(s): 2227 - 2241
Cited by: Papers (4)

Abstract HTML

The PartialSpoof Database and Countermeasures for the Detection of Short Fake Speech Segments Embedded in an Utterance
Lin Zhang; Xin Wang; Erica Cooper; Nicholas Evans; Junichi Yamagishi
Publication Year: 2023 , Page(s): 813 - 825
Cited by: Papers (32)

Abstract HTML

Audio Super-Resolution With Robust Speech Representation Learning of Masked Autoencoder
Seung-Bin Kim; Sang-Hoon Lee; Ha-Yeong Choi; Seong-Whan Lee
Publication Year: 2024 , Page(s): 1012 - 1022
Cited by: Papers (7)

Abstract HTML

BYOL for Audio: Exploring Pre-Trained General-Purpose Audio Representations
Daisuke Niizumi; Daiki Takeuchi; Yasunori Ohishi; Noboru Harada; Kunio Kashino
Publication Year: 2023 , Page(s): 137 - 151
Cited by: Papers (23)

Abstract HTML

Deep Learning-Based Non-Intrusive Multi-Objective Speech Assessment Model With Cross-Domain Features
Ryandhimas E. Zezario; Szu-Wei Fu; Fei Chen; Chiou-Shann Fuh; Hsin-Min Wang; Yu Tsao
Publication Year: 2023 , Page(s): 54 - 70
Cited by: Papers (45)

Abstract HTML

The LOCATA Challenge: Acoustic Source Localization and Tracking
Christine Evers; Heinrich W. Löllmann; Heinrich Mellmann; Alexander Schmidt; Hendrik Barfuss;

Patrick A. Naylor; Walter Kellermann
Publication Year: 2020 , Page(s): 1620 - 1643
Cited by: Papers (115)

Abstract HTML

Convolutional Neural Networks for Speech Recognition
Ossama Abdel-Hamid; Abdel-rahman Mohamed; Hui Jiang; Li Deng; Gerald Penn; Dong Yu
Publication Year: 2014 , Page(s): 1533 - 1545
Cited by: Papers (1604) | Patents (12)

Abstract HTML

Supervised Speech Separation Based on Deep Learning: An Overview
DeLiang Wang; Jitong Chen
Publication Year: 2018 , Page(s): 1702 - 1726
Cited by: Papers (997)

Abstract HTML

Anti-Spoofing for Text-Independent Speaker Verification: An Initial Database, Comparison of Countermeasures, and Human Performance
Zhizheng Wu; Phillip L. De Leon; Cenk Demiroglu; Ali Khodabakhsh; Simon King; Zhen-Hua Ling; Daisuke Saito; Bryan Stewart; Tomoki Toda; Mirjam Wester; Junichi Yamagishi
Publication Year: 2016 , Page(s): 768 - 783
Cited by: Papers (68) | Patents (24)

Abstract HTML

Enhancing Conformer-Based Sound Event Detection Using Frequency Dynamic Convolutions and BEATs Audio Embeddings
Sara Barahona; Diego de Benito-Gorrón; Doroteo T. Toledano; Daniel Ramos
Publication Year: 2024 , Page(s): 3896 - 3907

Abstract HTML

Analyzing the Robustness of Vision & Language Models
Alexander Shirnin; Nikita Andreev; Sofia Potapova; Ekaterina Artemova
Publication Year: 2024 , Page(s): 2751 - 2763

Abstract HTML

Binaural Sound Source Distance Estimation and Localization for a Moving Listener
Daniel Aleksander Krause; Guillermo García-Barrios; Archontis Politis; Annamaria Mesaros
Publication Year: 2024 , Page(s): 996 - 1011
Cited by: Papers (6)

Abstract HTML

An Overview of Deep-Learning-Based Audio-Visual Speech Enhancement and Separation
Daniel Michelsanti; Zheng-Hua Tan; Shi-Xiong Zhang; Yong Xu; Meng Yu; Dong Yu; Jesper Jensen
Publication Year: 2021 , Page(s): 1368 - 1396
Cited by: Papers (167)

Abstract HTML

Bayesian Parameter-Efficient Fine-Tuning for Overcoming Catastrophic Forgetting
Haolin Chen; Philip N. Garner
Publication Year: 2024 , Page(s): 4253 - 4262

Abstract HTML

Decoupling Speaker-Independent Emotions for Voice Conversion via Source-Filter Networks
Zhaojie Luo; Shoufeng Lin; Rui Liu; Jun Baba; Yuichiro Yoshikawa; Hiroshi Ishiguro
Publication Year: 2023 , Page(s): 11 - 24
Cited by: Papers (6)

Abstract HTML

Cooperative Scene-Event Modelling for Acoustic Scene Classification
Yuanbo Hou; Bo Kang; Andrew Mitchell; Wenwu Wang; Jian Kang; Dick Botteldooren
Publication Year: 2024 , Page(s): 68 - 82
Cited by: Papers (6)

Abstract HTML

Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence
Matteo Torcoli; Thorsten Kastner; Jürgen Herre
Publication Year: 2021 , Page(s): 1530 - 1541
Cited by: Papers (41)

Abstract HTML

NeuroHeed: Neuro-Steered Speaker Extraction Using EEG Signals
Zexu Pan; Marvin Borsdorf; Siqi Cai; Tanja Schultz; Haizhou Li
Publication Year: 2024 , Page(s): 4456 - 4470
Cited by: Papers (7)

Abstract HTML

Selective Acoustic Feature Enhancement for Speech Emotion Recognition With Noisy Speech
Seong-Gyun Leem; Daniel Fulford; Jukka-Pekka Onnela; David Gard; Carlos Busso
Publication Year: 2024 , Page(s): 917 - 929
Cited by: Papers (6)

Abstract HTML

AudioLM: A Language Modeling Approach to Audio Generation
Zalán Borsos; Raphaël Marinier; Damien Vincent; Eugene Kharitonov; Olivier Pietquin; Matt Sharifi; Dominik Roblek; Olivier Teboul; David Grangier; Marco Tagliasacchi; Neil Zeghidour
Publication Year: 2023 , Page(s): 2523 - 2533
Cited by: Papers (166)

Abstract HTML

Audio-Visual Cross-Attention Network for Robotic Speaker Tracking
Xinyuan Qian; Zhengdong Wang; Jiadong Wang; Guohui Guan; Haizhou Li
Publication Year: 2023 , Page(s): 550 - 562
Cited by: Papers (20)

Abstract HTML

Investigating the Design Space of Diffusion Models for Speech Enhancement
Philippe Gonzalez; Zheng-Hua Tan; Jan Østergaard; Jesper Jensen; Tommy Sonne Alstrøm; Tobias May
Publication Year: 2024 , Page(s): 4486 - 4500
Cited by: Papers (2)

Abstract HTML

Wavelet Multiresolution Analysis Based Speech Emotion Recognition System Using 1D CNN LSTM Networks
Aditya Dutt; Paul Gader
Publication Year: 2023 , Page(s): 2043 - 2054
Cited by: Papers (11)

Abstract HTML

Determined Blind Source Separation Unifying Independent Vector Analysis and Nonnegative Matrix Factorization
Daichi Kitamura; Nobutaka Ono; Hiroshi Sawada; Hirokazu Kameoka; Hiroshi Saruwatari
Publication Year: 2016 , Page(s): 1626 - 1641

Cited by: Papers (249) | Patents (2)

Abstract HTML

Real-Time Multichannel Deep Speech Enhancement in Hearing Aids: Comparing Monaural and Binaural Processing in Complex Acoustic Scenarios
Nils L. Westhausen; Hendrik Kayser; Theresa Jansen; Bernd T. Meyer
Publication Year: 2024 , Page(s): 4596 - 4606
Cited by: Papers (1)

Abstract HTML

Sound Field Estimation Based on Physics-Constrained Kernel Interpolation Adapted to Environment
Juliano G. C. Ribeiro; Shoichi Koyama; Ryosuke Horiuchi; Hiroshi Saruwatari
Publication Year: 2024 , Page(s): 4369 - 4383
Cited by: Papers (5)

Abstract HTML

Overview of Speaker Modeling and Its Applications: From the Lens of Deep Speaker Representation Learning
Shuai Wang; Zhengyang Chen; Kong Aik Lee; Yanmin Qian; Haizhou Li
Publication Year: 2024 , Page(s): 4971 - 4998
Cited by: Papers (1)

Abstract HTML

Overview and Evaluation of Sound Event Localization and Detection in DCASE 2019
Archontis Politis; Annamaria Mesaros; Sharath Adavanne; Toni Heittola; Tuomas Virtanen
Publication Year: 2021 , Page(s): 684 - 698
Cited by: Papers (75)

Abstract HTML

From Raw Speech to Fixed Representations: A Comprehensive Evaluation of Speech Embedding Techniques
Dejan Porjazovski; Tamás Grósz; Mikko Kurimo
Publication Year: 2024 , Page(s): 3546 - 3560

Abstract HTML

EfficientTTS 2: Variational End-to-End Text-to-Speech Synthesis and Voice Conversion
Chenfeng Miao; Qingying Zhu; Minchuan Chen; Jun Ma; Shaojun Wang; Jing Xiao
Publication Year: 2024 , Page(s): 1650 - 1661
Cited by: Papers (6)

Abstract HTML

Spatial Active Noise Control Based on Kernel Interpolation of Sound Field
Shoichi Koyama; Jesper Brunnström; Hayato Ito; Natsuki Ueno; Hiroshi Saruwatari
Publication Year: 2021 , Page(s): 3052 - 3063
Cited by: Papers (25)

Abstract HTML

ASVspoof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild
Xuechen Liu; Xin Wang; Md Sahidullah; Jose Patino; Héctor Delgado; Tomi Kinnunen; Massimiliano Todisco; Junichi Yamagishi; Nicholas Evans; Andreas Nautsch; Kong Aik Lee
Publication Year: 2023 , Page(s): 2507 - 2522
Cited by: Papers (84)

Abstract HTML

Submit Manuscript

Submission Guidelines

Become a Reviewer

Open Access Publishing Options

Add Title To My Alerts Add to My Favorites

Contact Information
Editor-in-Chief

Paris Smaragdis
University of Illinois at Urbana-Champaign
Champaign
USA
paris@illinois.edu

Editorial Board

Author Center

IEEE/ACM Transactions on Audio, Speech, and Language Processing

Additional Information

Information for Authors

Publishing Policies