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 - September 2024

Popular Article Feed

Includes the 50 most frequently accessed documents for this publication.

Export | Email Selected Results

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

Date
September 2024

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. **Joint Dual Learning With Mutual Information Maximization for Natural Language Understanding and Generation in Dialogues**
Shang-Yu Su; Yung-Sung Chung; Yun-Nung Chen
Publication Year: 2024 , Page(s): 2445 - 2452
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. **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

6. **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

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

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

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

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

Accented Text-to-Speech Synthesis With Limited Data
Xuehao Zhou; Mingyang Zhang; Yi Zhou; Zhizheng Wu; Haizhou Li
Publication Year: 2024 , Page(s): 1699 - 1711
Cited by: Papers (3)
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

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

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

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

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

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

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

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

Golden Gemini is All You Need: Finding the Sweet Spots for Speaker Verification
Tianchi Liu; Kong Aik Lee; Qiongqiong Wang; Haizhou Li
Publication Year: 2024 , Page(s): 2324 - 2337
Cited by: Papers (6)

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

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

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

WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research
Xinhao Mei; Chutong Meng; Haohe Liu; Qiuqiang Kong; Tom Ko; Chengqi Zhao; Mark D. Plumbley; Yuexian Zou; Wenwu Wang

Publication Year: 2024 , Page(s): 3339 - 3354
Cited by: Papers (51)

Abstract HTML

Diffsound: Discrete Diffusion Model for Text-to-Sound Generation
Dongchao Yang; Jianwei Yu; Helin Wang; Wen Wang; Chao Weng; Yuexian Zou; Dong Yu
Publication Year: 2023 , Page(s): 1720 - 1733
Cited by: Papers (80)

Abstract HTML

SpeechPrompt: Prompting Speech Language Models for Speech Processing Tasks
Kai-Wei Chang; Haibin Wu; Yu-Kai Wang; Yuan-Kuei Wu; Hua Shen; Wei-Cheng Tseng; Iu-Thing Kang; Shang-Wen Li; Hung-Yi Lee
Publication Year: 2024 , Page(s): 3730 - 3744
Cited by: Papers (2)

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

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

SpeechX: Neural Codec Language Model as a Versatile Speech Transformer
Xiaofei Wang; Manthan Thakker; Zhuo Chen; Naoyuki Kanda; Sefik Emre Eskimez; Sanyuan Chen; Min Tang; Shujie Liu; Jinyu Li; Takuya Yoshioka
Publication Year: 2024 , Page(s): 3355 - 3364
Cited by: Papers (17)

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

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

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

Unsupervised Music Source Separation Using Differentiable Parametric Source Models

Kilian Schulze-Forster; Gaël Richard; Liam Kelley; Clement S. J. Doire; Roland Badeau
Publication Year: 2023 , Page(s): 1276 - 1289
Cited by: Papers (12)

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

Spatial Analysis and Synthesis Methods: Subjective and Objective Evaluations Using Various Microphone Arrays in the Auralization of a Critical Listening Room
Alan Pawlak; Hyunkook Lee; Aki Mäkivirta; Thomas Lund
Publication Year: 2024 , Page(s): 3986 - 4001

Abstract HTML

InstructTTS: Modelling Expressive TTS in Discrete Latent Space With Natural Language Style Prompt
Dongchao Yang; Songxiang Liu; Rongjie Huang; Chao Weng; Helen Meng
Publication Year: 2024 , Page(s): 2913 - 2925
Cited by: Papers (20)

Abstract HTML

Music Source Separation With Band-Split RNN
Yi Luo; Jianwei Yu
Publication Year: 2023 , Page(s): 1893 - 1901
Cited by: Papers (63)

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

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

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

Speaker Distance Estimation in Enclosures From Single-Channel Audio
Michael Neri; Archontis Politis; Daniel Aleksander Krause; Marco Carli; Tuomas Virtanen
Publication Year: 2024 , Page(s): 2242 - 2254
Cited by: Papers (4)

Abstract HTML

Machine Speech Chain
Andros Tjandra; Sakriani Sakti; Satoshi Nakamura
Publication Year: 2020 , Page(s): 976 - 989

Cited by: Papers (19)

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

Deep Prior-Based Audio Inpainting Using Multi-Resolution Harmonic Convolutional Neural Networks
Federico Miotello; Mirco Pezzoli; Luca Comanducci; Fabio Antonacci; Augusto Sarti
Publication Year: 2024 , Page(s): 113 - 123
Cited by: Papers (7)

Abstract HTML

Multi-Channel Conversational Speaker Separation via Neural Diarization
Hassan Taherian; DeLiang Wang
Publication Year: 2024 , Page(s): 2467 - 2476
Cited by: Papers (7)

Abstract HTML

Convergence and Performance Analysis of Classical, Hybrid, and Deep Acoustic Echo Control
Ernst Seidel; Pejman Mowlaee; Tim Fingscheidt
Publication Year: 2024 , Page(s): 2857 - 2870
Cited by: Papers (2)

Abstract HTML

Textless Unit-to-Unit Training for Many-to-Many Multilingual Speech-to-Speech Translation
Minsu Kim; Jeongsoo Choi; Dahun Kim; Yong Man Ro
Publication Year: 2024 , Page(s): 3934 - 3946
Cited by: Papers (4)

Abstract HTML

TF-GridNet: Integrating Full- and Sub-Band Modeling for Speech Separation
Zhong-Qiu Wang; Samuele Cornell; Shukjae Choi; Younglo Lee; Byeong-Yeol Kim; Shinji Watanabe
Publication Year: 2023 , Page(s): 3221 - 3236
Cited by: Papers (67)

Abstract HTML

On the Generalization Ability of Complex-Valued Variational U-Networks for Single-Channel Speech Enhancement
Eike J. Nustede; Jörn Anemüller
Publication Year: 2024 , Page(s): 3838 - 3849

Abstract HTML

StoRM: A Diffusion-Based Stochastic Regeneration Model for Speech Enhancement and Dereverberation
Jean-Marie Lemercier; Julius Richter; Simon Welker; Timo Gerkmann
Publication Year: 2023 , Page(s): 2724 - 2737
Cited by: Papers (46)

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