Shinji Watanabe

List of Publications by Year
in descending order

274 papers	10,651 citations	331259 21 h-index	182168 51 g-index
278 all docs	278 docs citations	278 times ranked	3322 citing authors

#	Article	IF	Citations
1	Arabic speech recognition by end-to-end, modular systems and human. Computer Speech and Language, 2022, 71, 101272.	2.9	18
2	A review of speaker diarization: Recent advances with deep learning. Computer Speech and Language, 2022, 72, 101317.	2.9	90
3	Joint speaker diarization and speech recognition based on region proposal networks. Computer Speech and Language, 2022, 72, 101316.	2.9	3
4	Deep learning based multi-source localization with source splitting and its effectiveness in multi-talker speech recognition. Computer Speech and Language, 2022, 75, 101360.	2.9	38
5	Encoder-Decoder Based Attractors for End-to-End Neural Diarization. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 1493-1507.	4.0	19
6	Train from scratch: Single-stage joint training of speech separation and recognition. Computer Speech and Language, 2022, 76, 101387.	2.9	3
7	Conditional Diffusion Probabilistic Model for Speech Enhancement. , 2022, , .		36
8	Investigating Self-Supervised Learning for Speech Enhancement and Separation. , 2022, , .		20
9	The First Multimodal Information Based Speech Processing (Misp) Challenge: Data, Tasks, Baselines And Results. , 2022, , .		16
10	Extended Graph Temporal Classification for Multi-Speaker End-to-End ASR. , 2022, , .		1
11	Torchaudio: Building Blocks for Audio and Speech Processing. , 2022, , .		29
12	Improving Non-Autoregressive End-to-End Speech Recognition with Pre-Trained Acoustic and Language Models. , 2022, , .		8
13	Sequence Transduction with Graph-Based Supervision. , 2022, , .		4
14	Non-Autoregressive End-To-End Automatic Speech Recognition Incorporating Downstream Natural Language Processing. , 2022, , .		3
15	Joint Modeling of Code-Switched and Monolingual ASR via Conditional Factorization. , 2022, , .		4
16	ESPnet-SLU: Advancing Spoken Language Understanding Through ESPnet. , 2022, , .		20
17	Joint Speech Recognition and Audio Captioning. , 2022, , .		3
18	Towards Low-Distortion Multi-Channel Speech Enhancement: The ESPNET-Se Submission to the L3DAS22 Challenge. , 2022, , .		10

#	Article	IF	Citations
19	Integrating Multiple ASR Systems into NLP Backend with Attention Fusion. , 2022, , .		2
20	Multi-Channel End-To-End Neural Diarization with Distributed Microphones. , 2022, , .		4
21	SRU++: Pioneering Fast Recurrence with Attention for Speech Recognition. , 2022, , .		4
22	An Exploration of Hubert with Large Number of Cluster Units and Model Assessment Using Bayesian Information Criterion. , 2022, , .		1
23	Run-and-Back Stitch Search: Novel Block Synchronous Decoding For Streaming Encoder-Decoder ASR. , 2022, , .		0
24	A dilemma of ground truth in noisy speech separation and an approach to lessen the impact of imperfect training data. Computer Speech and Language, 2022, , 101410.	2.9	0
25	Improving Frame-Online Neural Speech Enhancement With Overlapped-Frame Prediction. IEEE Signal Processing Letters, 2022, 29, 1422-1426.	2.1	0
26	Far-Field Automatic Speech Recognition. Proceedings of the IEEE, 2021, 109, 124-148.	16.4	54
27	End-to-End Speaker Diarization Conditioned on Speech Activity and Overlap Detection. , 2021, , .		8
28	DOVER-Lap: A Method for Combining Overlap-Aware Diarization Outputs. , 2021, , .		19
29	Integration of Speech Separation, Diarization, and Recognition for Multi-Speaker Meetings: System Description, Comparison, and Analysis. , 2021, , .		27
30	Non-Autoregressive Transformer for Speech Recognition. IEEE Signal Processing Letters, 2021, 28, 121-125.	2.1	46
31	Online End-To-End Neural Diarization with Speaker-Tracing Buffer. , 2021, , .		12
32	Dual-Path Modeling for Long Recording Speech Separation in Meetings. , 2021, , .		5
33	Gaussian Kernelized Self-Attention for Long Sequence Data and its Application to CTC-Based Speech Recognition. , 2021, , .		2
34	Improving RNN Transducer with Target Speaker Extraction and Neural Uncertainty Estimation. , 2021, , .		2
35	Recent Developments on Espnet Toolkit Boosted By Conformer. , 2021, , .		80
36	ORTHROS: non-autoregressive end-to-end speech translation With dual-decoder. , 2021, , .		10

#	Article	IF	Citations
37	End-to-End Dereverberation, Beamforming, and Speech Recognition with Improved Numerical Stability and Advanced Frontend. , 2021, , .		20
38	Intermediate Loss Regularization for CTC-Based Speech Recognition. , 2021, , .		47
39	Eat: Enhanced ASR-TTS for Self-Supervised Speech Recognition. , 2021, , .		12
40	The 2020 ESPnet Update: New Features, Broadened Applications, Performance Improvements, and Future Plans. , 2021, , .		16
41	Improved Mask-CTC for Non-Autoregressive End-to-End ASR. , 2021, , .		27
42	Directional ASR: A New Paradigm for E2E Multi-Speaker Speech Recognition with Source Localization. , 2021, , .		6
43	Training Noisy Single-Channel Speech Separation with Noisy Oracle Sources: A Large Gap and a Small Step. , 2021, , .		2
44	End-To-End Diarization for Variable Number of Speakers with Local-Global Networks and Discriminative Speaker Embeddings. , 2021, , .		14
45	End-To-End Speaker Diarization as Post-Processing. , 2021, , .		13
46	ESPnet-SE: End-To-End Speech Enhancement and Separation Toolkit Designed for ASR Integration. , 2021, , .		41
47	Dual-Path RNN for Long Recording Speech Separation. , 2021, , .		9
48	Sequential Multi-Frame Neural Beamforming for Speech Separation and Enhancement. , 2021, , .		21
49	Streaming Transformer Asr With Blockwise Synchronous Beam Search. , 2021, , .		17
50	An investigation of neural uncertainty estimation for target speaker extraction equipped RNN transducer. Computer Speech and Language, 2021, 73, 101327.	2.9	0
51	Closing the Gap Between Time-Domain Multi-Channel Speech Enhancement on Real and Simulation Conditions. , 2021, , .		8
52	Attention-Based Multi-Hypothesis Fusion for Speech Summarization. , 2021, , .		7
53	Cross-Lingual Transfer for Speech Processing Using Acoustic Language Similarity. , 2021, , .		0
54	Towards Neural Diarization for Unlimited Numbers of Speakers Using Global and Local Attractors. , 2021, , .		13

#	Article	IF	Citations
55	Conferencingspeech Challenge: Towards Far-Field Multi-Channel Speech Enhancement for Video Conferencing. , 2021, , .		10
56	A Comparative Study on Non-Autoregressive Modelings for Speech-to-Text Generation. , 2021, , .		23
57	On Prosody Modeling for ASR+TTS Based Voice Conversion. , 2021, , .		5
58	An Exploration of Self-Supervised Pretrained Representations for End-to-End Speech Recognition. , 2021, , .		26
59	Fast-MD: Fast Multi-Decoder End-to-End Speech Translation with Non-Autoregressive Hidden Intermediates. , 2021, , .		4
60	A Study of Transducer Based End-to-End ASR with ESPnet: Architecture, Auxiliary Loss and Decoding Strategies. , 2021, , .		10
61	Multi-Stream End-to-End Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 646-655.	4.0	10
62	Espnet-TTS: Unified, Reproducible, and Integratable Open Source End-to-End Text-to-Speech Toolkit. , 2020, , .		82
63	Attention-Based ASR with Lightweight and Dynamic Convolutions. , 2020, , .		13
64	Semi-Supervised Speaker Adaptation for End-to-End Speech Synthesis with Pretrained Models. , 2020, , .		8
65	A Practical Two-Stage Training Strategy for Multi-Stream End-to-End Speech Recognition. , 2020, , .		4
66	End-to-End Automatic Speech Recognition Integrated with CTC-Based Voice Activity Detection. , 2020, , .		20
67	Speaker Diarization with Region Proposal Network. , 2020, , .		27
68	Weakly-Supervised Sound Event Detection with Self-Attention. , 2020, , .		28
69	End-To-End Multi-Speaker Speech Recognition With Transformer. , 2020, , .		49
70	Far-Field Location Guided Target Speech Extraction Using End-to-End Speech Recognition Objectives. , 2020, , .		21
71	Improving End-to-End Single-Channel Multi-Talker Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 1385-1394.	4.0	16
72	Automated Development of DNN Based Spoken Language Systems Using Evolutionary Algorithms. Natural Computing Series, 2020, , 97-129.	2.2	2

#	Article	IF	Citations
73	ESPnet-ST: All-in-One Speech Translation Toolkit. , 2020, , .		53
74	Acoustic Modeling for Distant Multi-talker Speech Recognition with Single- and Multi-channel Branches. , 2019, , .		23
75	Language Model Integration Based on Memory Control for Sequence to Sequence Speech Recognition. , 2019, , .		1
76	End-to-end Monaural Multi-speaker ASR System without Pretraining. , 2019, , .		39
77	Introduction to the Issue on Far-Field Speech Processing in the Era of Deep Learning: Speech Enhancement, Separation, and Recognition. IEEE Journal on Selected Topics in Signal Processing, 2019, 13, 785-786.	7.3	2
78	Transfer Learning of Language-independent End-to-end ASR with Language Model Fusion. , 2019, , .		23
79	Speech Processing for Digital Home Assistants: Combining Signal Processing With Deep-Learning Techniques. IEEE Signal Processing Magazine, 2019, 36, 111-124.	4.6	100
80	Stream Attention-based Multi-array End-to-end Speech Recognition. , 2019, , .		10
81	Improving End-to-end Speech Recognition with Pronunciation-assisted Sub-word Modeling. , 2019, , .		24
82	Promising Accurate Prefix Boosting for Sequence-to-sequence ASR. , 2019, , .		6
83	Cycle-consistency Training for End-to-end Speech Recognition. , 2019, , .		41
84	Phasebook and Friends: Leveraging Discrete Representations for Source Separation. IEEE Journal on Selected Topics in Signal Processing, 2019, 13, 370-382.	7.3	44
85	Joint Acoustic and Class Inference for Weakly Supervised Sound Event Detection. , 2019, , .		8
86	The Phasebook: Building Complex Masks via Discrete Representations for Source Separation. , 2019, , .		7
87	Acoustic Modeling for Overlapping Speech Recognition: Jhu Chime-5 Challenge System. , 2019, , .		15
88	Semi-supervised End-to-end Speech Recognition Using Text-to-speech and Autoencoders. , 2019, , .		29
89	Espresso: A Fast End-to-End Neural Speech Recognition Toolkit. , 2019, , .		32
90	Transformer ASR with Contextual Block Processing. , 2019, , .		23

#	Article	IF	Citations
91	Generalized Weighted-Prediction-Error Dereverberation with Varying Source Priors For Reverberant Speech Recognition. , 2019, , .		4
92	MIMO-Speech: End-to-End Multi-Channel Multi-Speaker Speech Recognition. , 2019, , .		61
93	Multilingual End-to-End Speech Translation. , 2019, , .		37
94	End-to-End Neural Speaker Diarization with Self-Attention. , 2019, , .		91
95	Simultaneous Speech Recognition and Speaker Diarization for Monaural Dialogue Recordings with Target-Speaker Acoustic Models. , 2019, , .		23
96	A Comparative Study on Transformer vs RNN in Speech Applications. , 2019, , .		344
97	Using ASR Methods for OCR. , 2019, , .		14
98	Weakly-Supervised Deep Recurrent Neural Networks for Basic Dance Step Generation. , 2019, , .		28
99	Analysis of Robustness of Deep Single-Channel Speech Separation Using Corpora Constructed From Multiple Domains. , 2019, , .		7
100	Speech Enhancement Using End-to-End Speech Recognition Objectives. , 2019, , .		32
101	CNN-based Multichannel End-to-End Speech Recognition for Everyday Home Environments. , 2019, , .		8
102	Preparation of monodisperse fully aromatic polyimide particles via the polycondensation of diethyl hexafluoroisopropylidenediphthalate with 4,4â€²-diaminodiphenylether in ethylene glycol. Polymer Journal, 2019, 51, 405-412.	1.3	4
103	Evolution-Strategy-Based Automation of System Development for High-Performance Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2019, 27, 77-88.	4.0	11
104	Model parameter learning using Kullbackâ€"Leibler divergence. Physica A: Statistical Mechanics and Its Applications, 2018, 491, 549-559.	1.2	1
105	Back-Translation-Style Data Augmentation for end-to-end ASR. , 2018, , .		49
106	Speaker Adaptation for Multichannel End-to-End Speech Recognition. , 2018, , .		31
107	Multilingual Sequence-to-Sequence Speech Recognition: Architecture, Transfer Learning, and Language Modeling. , 2018, , .		55
108	End-to-end Speech Recognition With Word-Based Rnn Language Models. , 2018, , .		64

#	Article	IF	Citations
109	End-to-End Multi-Speaker Speech Recognition. , 2018, , .		44
110	Low-Resource Contextual Topic Identification on Speech. , 2018, , .		0
111	An End-to-End Language-Tracking Speech Recognizer for Mixed-Language Speech. , 2018, , .		33
112	A Purely End-to-End System for Multi-speaker Speech Recognition. , 2018, , .		42
113	Multi-microphone speech recognition integrating beamforming, robust feature extraction, and advanced DNN/RNN backend. Computer Speech and Language, 2017, 46, 401-418.	2.9	23
114	Deep long short-term memory adaptive beamforming networks for multichannel robust speech recognition. , 2017, , .		52
115	BLSTM-HMM hybrid system combined with sound activity detection network for polyphonic Sound Event Detection. , 2017, , .		14
116	The third â€˜CHiMEâ€™ speech separation and recognition challenge: Analysis and outcomes. Computer Speech and Language, 2017, 46, 605-626.	2.9	70
117	An analysis of environment, microphone and data simulation mismatches in robust speech recognition. Computer Speech and Language, 2017, 46, 535-557.	2.9	202
118	Duration-Controlled LSTM for Polyphonic Sound Event Detection. IEEE/ACM Transactions on Audio Speech and Language Processing, 2017, 25, 2059-2070.	4.0	63
119	Student-teacher network learning with enhanced features. , 2017, , .		45
120	Hybrid CTC/Attention Architecture for End-to-End Speech Recognition. IEEE Journal on Selected Topics in Signal Processing, 2017, 11, 1240-1253.	7.3	405
121	Deep Recurrent Networks for Separation and Recognition of Single-Channel Speech in Nonstationary Background Audio. , 2017, , 165-186.		23
122	The CHiME Challenges: Robust Speech Recognition in Everyday Environments. , 2017, , 327-344.		16
123	Discriminative Beamforming with Phase-Aware Neural Networks for Speech Enhancement and Recognition. , 2017, , 79-104.		0
124	Novel Deep Architectures in Speech Processing. , 2017, , 135-164.		2
125	Prior-based Binary Masking and Discriminative Methods for Reverberant and Noisy Speech Recognition Using Distant Stereo Microphones. Journal of Information Processing, 2017, 25, 407-416.	0.3	0
126	Unified Architecture for Multichannel End-to-End Speech Recognition With Neural Beamforming. IEEE Journal on Selected Topics in Signal Processing, 2017, 11, 1274-1288.	7.3	54

#	Article	IF	Citations
127	Joint CTC-attention based end-to-end speech recognition using multi-task learning. , 2017, , .		449
128	Language independent end-to-end architecture for joint language identification and speech recognition. , 2017, , .		66
129	Multi-level language modeling and decoding for open vocabulary end-to-end speech recognition. , 2017, , .		31
130	Does speech enhancement work with end-to-end ASR objectives?: Experimental analysis of multichannel end-to-end ASR. , 2017, , .		4
131	Joint CTC/attention decoding for end-to-end speech recognition. , 2017, , .		73
132	Nested Gibbs sampling for mixture-of-mixture model and its application to speaker clustering. APSIPA Transactions on Signal and Information Processing, 2016, 5, .	2.6	1
133	Dialog state tracking with attention-based sequence-to-sequence learning. , 2016, , .		12
134	Automated structure discovery and parameter tuning of neural network language model based on evolution strategy. , 2016, , .		11
135	Beamforming networks using spatial covariance features for far-field speech recognition. , 2016, , .		11
136	Minimum word error training of long short-term memory recurrent neural network language models for speech recognition. , 2016, , .		14
137	Driver confusion status detection using recurrent neural networks. , 2016, , .		8
138	Deep clustering: Discriminative embeddings for segmentation and separation. , 2016, , .		732
139	Deep beamforming networks for multi-channel speech recognition. , 2016, , .		111
140	Sequence summarizing neural network for speaker adaptation. , 2016, , .		46
141	High-accuracy user identification using EEG biometrics. , 2016, 2016, 854-858.		36
142	Deep unfolding for multichannel source separation. , 2016, , .		18
143	A sampling-based speaker clustering using utterance-oriented Dirichlet process mixture model and its evaluation on large-scale data. APSIPA Transactions on Signal and Information Processing, 2015, 4, .	2.6	3
144	Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks. , 2015, , .		414

#	Article	IF	Citations
145	Feature-space structural MAPLR with regression tree-based multiple transformation matrices for DNN. , 2015, , .		3
146	Effectiveness of dereverberation, feature transformation, discriminative training methods, and system combination approach for various reverberant environments. Eurasip Journal on Advances in Signal Processing, 2015, 2015, .	1.0	0
147	Rutheniumâ€Catalyzed Dehydrogenative Aromatic Cï£¿H Silylation of Benzamides with Hydrosilanes. Advanced Synthesis and Catalysis, 2015, 357, 2229-2232.	2.1	21
148	Living Cationic Polymerization of Vinyl Ether with a Thienyl Group. Kobunshi Ronbunshu, 2015, 72, 433-439.	0.2	2
149	Preparation of aliphaticâ€"aromatic polyimide particles by polycondensation of diethyl hexafluoroisopropylidenediphthalate and diaminooctane in ethylene glycol. High Performance Polymers, 2015, 27, 183-190.	0.8	9
150	Robust speech recognition in unknown reverberant and noisy conditions. , 2015, , .		23
151	Automation of system building for state-of-the-art large vocabulary speech recognition using evolution strategy. , 2015, , .		18
152	The MERL/SRI system for the 3RD CHiME challenge using beamforming, robust feature extraction, and advanced speech recognition. , 2015, , .		28
153	The third â€˜CHiMEâ€™ speech separation and recognition challenge: Dataset, task and baselines. , 2015, , .		372
154	Discriminative method for recurrent neural network language models. , 2015, , .		11
155	Structure discovery of deep neural network based on evolutionary algorithms. , 2015, , .		46
156	Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR. Lecture Notes in Computer Science, 2015, , 91-99.	1.0	333
157	Log-linear dialog manager. , 2014, , .		0
158	Sequence discriminative training for low-rank deep neural networks. , 2014, , .		1
159	Deep recurrent de-noising auto-encoder and blind de-reverberation for reverberated speech recognition. , 2014, , .		56
160	Structural Bayesian Linear Regression for Hidden Markov Models. Journal of Signal Processing Systems, 2014, 74, 341-358.	1.4	11
161	Ensemble integration of calibrated speaker localization and statistical speech detection in domestic environments. , 2014, , .		5
162	Recurrent deep neural networks for robust speech recognition. , 2014, , .		93

#	Article	IF	Citations
163	Black box optimization for automatic speech recognition. , 2014, , .		20
164	Catalytic Activity for the Methane Steam Reforming Process Using Chemical Solution Deposition Derived Barium Titanate Hollow Particles with Perovskite Mono-phase. Journal of the Society of Powder Technology, Japan, 2014, 51, 337-342.	0.0	4
165	Training data selection with user's physical characteristics data for acceleration-based activity modeling. Personal and Ubiquitous Computing, 2013, 17, 451-463.	1.9	2
166	Speech recognition in living rooms: Integrated speech enhancement and recognition system based on spatial, spectral and temporal modeling of sounds. Computer Speech and Language, 2013, 27, 851-873.	2.9	17
167	Blocked Gibbs sampling based multi-scale mixture model for speaker clustering on noisy data. , 2013, , .		0
168	The second &#x2018;chime&#x2019; speech separation and recognition challenge: Datasets, tasks and baselines. , 2013, , .		177
169	Prior-shared feature and model space speaker adaptation by consistently employing map estimation. Speech Communication, 2013, 55, 415-431.	1.6	1
170	Cluster-based dynamic variance adaptation for interconnecting speech enhancement pre-processor and speech recognizer. Computer Speech and Language, 2013, 27, 350-368.	2.9	11
171	Influence relation estimation based on lexical entrainment in conversation. Speech Communication, 2013, 55, 329-339.	1.6	2
172	A generalized discriminative training framework for system combination. , 2013, , .		6
173	The second &#x2018;CHiME&#x2019; speech separation and recognition challenge: An overview of challenge systems and outcomes. , 2013, , .		48
174	Feature Enhancement With Joint Use of Consecutive Corrupted and Noise Feature Vectors With Discriminative Region Weighting. IEEE Transactions on Audio Speech and Language Processing, 2013, 21, 2172-2181.	3.8	6
175	Effectiveness of discriminative training and feature transformation for reverberated and noisy speech. , 2013, , .		12
176	Stereo-based feature enhancement using dictionary learning. , 2013, , .		0
177	Ensemble learning for speech enhancement. , 2013, , .		13
178	Preparation of barium titanate hollow particle by two-step chemical solution deposition. Journal of the Ceramic Society of Japan, 2013, 121, 80-83.	0.5	10
179	Handling uncertain observations in unsupervised topic-mixture language model adaptation. , 2012, , .		1
180	Fully Bayesian inference of multi-mixture Gaussian model and its evaluation using speaker clustering. , 2012, , .		3

#	Article	IF	Citations
181	Effect of dialog acts on word use in polylogue. , 2012, , .		0
182	Bag Of ARCS: New representation of speech segment features based on finite state machines. , 2012, , .		0
183	Discriminative feature transforms using differenced maximum mutual information. , 2012, , .		4
184	Noise suppression with unsupervised joint speaker adaptation and noise mixture model estimation. , 2012, , .		12
185	Structural Classification Methods Based on Weighted Finite-State Transducers for Automatic Speech Recognition. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 2240-2251.	3.8	6
186	Structured Discriminative Models For Speech Recognition: An Overview. IEEE Signal Processing Magazine, 2012, 29, 70-81.	4.6	22
187	MFCC enhancement using joint corrupted and noise feature space for highly non-stationary noise environments. , 2012, , .		8
188	Decoding network optimization using minimum transition error training. , 2012, , .		4
189	Basis vector orthogonalization for an improved kernel gradient matching pursuit method. , 2012, , .		1
190	Bayesian approaches to acoustic modeling: a review. APSIPA Transactions on Signal and Information Processing, 2012, 1, .	2.6	2
191	Ruthenium-catalyzed Ortho-selective Aromatic Câ€"H Silylation: Acceptorless Dehydrogenative Coupling of Hydrosilanes. Chemistry Letters, 2012, 41, 374-376.	0.7	49
192	Preparation of aliphatic polypyromellitimide particles by polycondensation of nylon-salt-type monomers derived from aliphatic diamines with diethyl pyromellitate in ethylene glycol. High Performance Polymers, 2012, 24, 710-716.	0.8	4
193	Frame-wise model re-estimation method based on Gaussian pruning with weight normalization for noise robust voice activity detection. Speech Communication, 2012, 54, 229-244.	1.6	10
194	Low-Latency Real-Time Meeting Recognition and Understanding Using Distant Microphones and Omni-Directional Camera. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 499-513.	3.8	65
195	Statistical Voice Conversion Based on Noisy Channel Model. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 1784-1794.	3.8	22
196	Discriminative approach to dynamic variance adaptation for noisy speech recognition. , 2011, , .		1
197	Unsupervised Activity Recognition with User's Physical Characteristics Data. , 2011, , .		58
198	Topic tracking language model for speech recognition. Computer Speech and Language, 2011, 25, 440-461.	2.9	22

#	Article	IF	Citations
199	High accurate model-integration-based voice conversion using dynamic features and model structure optimization. , 2011, , .		2
200	Subspace pursuit method for kernel-log-linear models. , 2011, , .		5
201	Bayesian linear regression for Hidden Markov Model based on optimizing variational bounds. , 2011, , .		9
202	Gibbs sampling based Multi-scale Mixture Model for speaker clustering. , 2011, , .		4
203	Non-stationary noise estimation method based on bias-residual component decomposition for robust speech recognition. , 2011, , .		2
204	Low-latency meeting recognition and understanding using distant microphones. , 2011, , .		2
205	Variance Compensation for Recognition of Reverberant Speech with Dereverberation Preprocessing. , 2011, , 225-255.		2
206	Preparation of monodisperse PMMA particles by dispersion polymerization of MMA using poly(styrene-co-methacrylic acid) copolymer as a steric stabilizer. Polymer Bulletin, 2010, 65, 543-550.	1.7	8
207	Predictorâ€"Corrector Adaptation by Using Time Evolution System With Macroscopic Time Scale. IEEE Transactions on Audio Speech and Language Processing, 2010, 18, 395-406.	3.8	3
208	A Sequential Pattern Classifier Based on Hidden Markov Kernel Machine and Its Application to Phoneme Classification. IEEE Journal on Selected Topics in Signal Processing, 2010, 4, 974-984.	7.3	8
209	Online Unsupervised Classification With Model Comparison in the Variational Bayes Framework for Voice Activity Detection. IEEE Journal on Selected Topics in Signal Processing, 2010, 4, 1071-1083.	7.3	12
210	Using online model comparison in the Variational Bayes framework for online unsupervised Voice Activity Detection. , 2010, , .		1
211	Search error risk minimization in Viterbi beam search for speech recognition. , 2010, , .		4
212	Online meeting recognizer with multichannel speaker diarization. , 2010, , .		8
213	A discriminative model for continuous speech recognition based on Weighted Finite State Transducers. , 2010, , .		3
214	Discriminative training based on an integrated view of MPE and MMI in margin and error space. , 2010, , .		31
215	Application of topic tracking model to language model adaptation and meeting analysis. , 2010, , .		0
216	Fast similarity search on a large speech data set with neighborhood graph indexing. , 2010, , .		5

#	Article	IF	Citations
217	Minimum Error Classification with geometric margin control. , 2010, , .		6
218	Real-time meeting recognition and understanding using distant microphones and omni-directional camera. , 2010, , .		13
219	On-line adaptation and Bayesian detection of environmental changes based on a macroscopic time evolution system. , 2009, , .		2
220	A unified view for discriminative objective functions based on negative exponential of difference measure between strings. , 2009, , .		6
221	Static and Dynamic Variance Compensation for Recognition of Reverberant Speech With Dereverberation Preprocessing. IEEE Transactions on Audio Speech and Language Processing, 2009, 17, 324-334.	3.8	50
222	Combined static and dynamic variance adaptation for efficient interconnection of speech enhancement pre-processor with speech recognizer. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	2
223	A unified interpretation of adaptation approaches based on a macroscopic time evolution system and indirect/direct adaptation approaches. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	2
224	Incremental Adaptation Based on a Macroscopic Time Evolution System. , 2007, , .		3
225	Preparation of Polystyrene-Polyimide Particles by Dispersion Polymerization of Styrene Using Poly(amic acid) as a Stabilizer. Polymer Journal, 2006, 38, 471-476.	1.3	9
226	Automatic determination of acoustic model topology using variational Bayesian estimation and clustering for large vocabulary continuous speech recognition. IEEE Transactions on Audio Speech and Language Processing, 2006, 14, 855-872.	3.8	12
227	Speech Recognition Based on Student's t-Distribution Derived from Total Bayesian Framework. IEICE Transactions on Information and Systems, 2006, E89-D, 970-980.	0.4	3
228	Selection of Shared-State Hidden Markov Model Structure Using Bayesian Criterion. IEICE Transactions on Information and Systems, 2005, E88-D, 1-9.	0.4	0
229	Variational Bayesian Estimation and Clustering for Speech Recognition. IEEE Transactions on Speech and Audio Processing, 2004, 12, 365-381.	2.0	79
230	Synthesis of Polymer Microspheres with Mercapto Groups by Polycondensation of 1,3-Propanedithiol and 1,6-Dibromohexane in the Presence of a Poly[styrene-alkylacrylamide] Latex. Polymer Journal, 2004, 36, 45-49.	1.3	1
231	SYNTHESIS OF BENZYLBORONATES VIA PALLADIUM-CATALYZED BORYLATION OF BENZYL HALIDES WITH PINACOLBORANE. Synthetic Communications, 2002, 32, 2513-2517.	1.1	26
232	Palladium-Catalyzed Borylation of Aryl Halides or Triflates with Dialkoxyborane:â€‰ A Novel and Facile Synthetic Route to Arylboronates. Journal of Organic Chemistry, 2000, 65, 164-168.	1.7	359
233	Solar neutrinos and leptonic CP violation. Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics, 1999, 468, 256-260.	1.5	31
234	Synthesis and characterization of new aromatic polyesters and a polyether derived from 2,2-bis(4-hydroxyphenyl)-1,2-diphenylethanone. Journal of Polymer Science Part A, 1998, 36, 2229-2235.	2.5	16

#	Article	IF	Citations
235	Synthesis of polymer microspheres with mercapto groups by polycondensation of 1,3-propanedithiol and 1,6-dibromohexane in the presence of a polystyrene latex. Macromolecular Rapid Communications, 1998, 19, 75-77.	2.0	3
236	Novel Palladium(0)-Catalyzed Coupling Reaction of Dialkoxyborane with Aryl Halides:â€‰ Convenient Synthetic Route to Arylboronates. Journal of Organic Chemistry, 1997, 62, 6458-6459.	1.7	297
237	Online Streaming End-to-End Neural Diarization Handling Overlapping Speech and Flexible Numbers of Speakers. , 0, , .		10
238	Single-Channel Multi-Speaker Separation Using Deep Clustering. , 0, , .		243
239	Improved MVDR Beamforming Using Single-Channel Mask Prediction Networks. , 0, , .		169
240	Advances in Joint CTC-Attention Based End-to-End Speech Recognition with a Deep CNN Encoder and RNN-LM. , 0, , .		165
241	Auxiliary Feature Based Adaptation of End-to-end ASR Systems. , 0, , .		26
242	ESPnet: End-to-End Speech Processing Toolkit. , 0, , .		612
243	Multi-Head Decoder for End-to-End Speech Recognition. , 0, , .		10
244	Semi-Supervised End-to-End Speech Recognition. , 0, , .		36
245	The Fifth 'CHiME' Speech Separation and Recognition Challenge: Dataset, Task and Baselines. , 0, , .		146
246	Diarization is Hard: Some Experiences and Lessons Learned for the JHU Team in the Inaugural DIHARD Challenge. , 0, , .		104
247	Student-Teacher Learning for BLSTM Mask-based Speech Enhancement. , 0, , .		14
248	Improving Transformer-Based End-to-End Speech Recognition with Connectionist Temporal Classification and Language Model Integration. , 0, , .		94
249	End-to-End Neural Speaker Diarization with Permutation-Free Objectives. , 0, , .		91
250	Study of the Performance of Automatic Speech Recognition Systems in Speakers with Parkinsonâ€™s Disease. , 0, , .		8
251	Semi-Supervised Sequence-to-Sequence ASR Using Unpaired Speech and Text. , 0, , .		35
252	Pre-Trained Text Embeddings for Enhanced Text-to-Speech Synthesis. , 0, , .		28

#	Article	IF	Citations
253	End-to-End Speaker Diarization for an Unknown Number of Speakers with Encoder-Decoder Based Attractors. , 0, , .		66
254	The Sequence-to-Sequence Baseline for the Voice Conversion Challenge 2020: Cascading ASR and TTS. , 0, , .		21
255	Probabilistic integration of joint density model and speaker model for voice conversion. , 0, , .		4
256	Large vocabulary continuous speech recognition using WFST-based linear classifier for structured data. , 0, , .		7
257	Speaker clustering based on utterance-oriented dirichlet process mixture model. , 0, , .		3
258	Fully Bayesian speaker clustering based on hierarchically structured utterance-oriented Dirichlet process mixture model. , 0, , .		2
259	Context-Sensitive and Role-Dependent Spoken Language Understanding Using Bidirectional and Attention LSTMs. , 0, , .		10
260	Data Selection by Sequence Summarizing Neural Network in Mismatch Condition Training. , 0, , .		3
261	Semi-Supervised Learning of a Pronunciation Dictionary from Disjoint Phonemic Transcripts and Text. , 0, , .		4
262	Coupled Initialization of Multi-Channel Non-Negative Matrix Factorization Based on Spatial and Spectral Information. , 0, , .		6
263	Multi-Modal Data Augmentation for End-to-end ASR. , 0, , .		32
264	Building State-of-the-art Distant Speech Recognition Using the CHiME-4 Challenge with a Setup of Speech Enhancement Baseline. , 0, , .		38
265	Effectiveness of Single-Channel BLSTM Enhancement for Language Identification. , 0, , .		0
266	Vectorized Beam Search for CTC-Attention-Based Speech Recognition. , 0, , .		18
267	End-to-End Multilingual Multi-Speaker Speech Recognition. , 0, , .		3
268	Analysis of Multilingual Sequence-to-Sequence Speech Recognition Systems. , 0, , .		7
269	Pretraining by Backtranslation for End-to-End ASR in Low-Resource Settings. , 0, , .		5
270	Auxiliary Interference Speaker Loss for Target-Speaker Speech Recognition. , 0, , .		9

#	Article	IF	Citations
271	End-to-End SpeakerBeam for Single Channel Target Speech Recognition. , 0, , .		7
272	Speaker Recognition Benchmark Using the CHiME-5 Corpus. , 0, , .		6
273	CHiME-6 Challenge: Tackling Multispeaker Speech Recognition for Unsegmented Recordings. , 0, , .		95
274	Mask CTC: Non-Autoregressive End-to-End ASR with CTC and Mask Predict. , 0, , .		60