Haizhou Li

List of Publications by Year
in descending order

528 papers	10,285 citations	109321 35 h-index	82547 72 g-index
533 all docs	533 docs citations	533 times ranked	4545 citing authors

#	Article	IF	Citations
1	A Tandem Learning Rule for Effective Training and Rapid Inference of Deep Spiking Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34, 446-460.	11.3	40
2	Emotion Intensity and its Control for Emotional Voice Conversion. IEEE Transactions on Affective Computing, 2023, 14, 31-48.	8.3	9
3	Progressive Tandem Learning for Pattern Recognition With Deep Spiking Neural Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44, 7824-7840.	13.9	33
4	Rectified Linear Postsynaptic Potential Function for Backpropagation in Deep Spiking Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33, 1947-1958.	11.3	55
5	EEG-Based Auditory Attention Detection via Frequency and Channel Neural Attention. IEEE Transactions on Human-Machine Systems, 2022, 52, 256-266.	3.5	8
6	Emotional voice conversion: Theory, databases and ESD. Speech Communication, 2022, 137, 1-18.	2.8	36
7	Neural Acoustic-Phonetic Approach for Speaker Verification With Phonetic Attention Mask. IEEE Signal Processing Letters, 2022, 29, 782-786.	3.6	7
8	STAnet: A Spatiotemporal Attention Network for Decoding Auditory Spatial Attention From EEG. IEEE Transactions on Biomedical Engineering, 2022, 69, 2233-2242.	4.2	28
9	Noise-robust voice conversion with domain adversarial training. Neural Networks, 2022, 148, 74-84.	5.9	9
10	Selective Listening by Synchronizing Speech With Lips. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 1650-1664.	5.8	9
11	Decoding Knowledge Transfer for Neural Text-to-Speech Training. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 1789-1802.	5.8	6
12	Self-Supervised Speaker Recognition with Loss-Gated Learning. , 2022, , .		17
13	L-SpEx: Localized Target Speaker Extraction. , 2022, , .		4
14	Visualtts: TTS with Accurate Lip-Speech Synchronization for Automatic Voice Over. , 2022, , .		3
15	A Hybrid Learning Framework for Deep Spiking Neural Networks with One-Spike Temporal Coding. , 2022, , .		1
16	Memobert: Pre-Training Model with Prompt-Based Learning for Multimodal Emotion Recognition. , 2022, , .		13
17	Genre-Conditioned Acoustic Models for Automatic Lyrics Transcription of Polyphonic Music. , 2022, , .		4
18	Speaker Extraction With Co-Speech Gestures Cue. IEEE Signal Processing Letters, 2022, 29, 1467-1471.	3.6	5

#	Article	IF	Citations
19	Optimizing the Perceptual Quality of Time-Domain Speech Enhancement with Reinforcement Learning. Tsinghua Science and Technology, 2022, 27, 939-947.	6.1	5
20	Deep Learning Approaches in Topics of Singing Information Processing. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 2422-2451.	5.8	7
21	An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 132-157.	5.8	123
22	Exploiting Morphological and Phonological Features to Improve Prosodic Phrasing for Mongolian Speech Synthesis. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 274-285.	5.8	19
23	Multi-Tone Phase Coding of Interaural Time Difference for Sound Source Localization With Spiking Neural Networks. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 2656-2670.	5.8	14
24	Optimizing Voice Conversion Network with Cycle Consistency Loss of Speaker Identity. , 2021, , .		8
25	Capsule Network based End-to-end System for Detection of Replay Attacks. , 2021, , .		4
26	An Automatic Sound Classification Framework with Non-volatile Memory. , 2021, , 415-438.		0
27	Vaw-Gan For Disentanglement And Recomposition Of Emotional Elements In Speech. , 2021, , .		11
28	D-Score: Holistic Dialogue Evaluation Without Reference. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 2502-2516.	5.8	5
29	Three-Dimensional Speaker Localization: Audio-Refined Visual Scaling Factor Estimation. IEEE Signal Processing Letters, 2021, 28, 1405-1409.	3.6	3
30	Representation Learning with Spectro-Temporal-Channel Attention for Speech Emotion Recognition. , 2021, , .		17
31	Multi-Stage Speaker Extraction with Utterance and Frame-Level Reference Signals. , 2021, , .		17
32	Muse: Multi-Modal Target Speaker Extraction with Visual Cues. , 2021, , .		17
33	Factorized WaveNet for voice conversion with limited data. Speech Communication, 2021, 130, 45-54.	2.8	4
34	Data Augmentation with Signal Companding for Detection of Logical Access Attacks. , 2021, , .		12
35	Graphspeech: Syntax-Aware Graph Attention Network for Neural Speech Synthesis. , 2021, , .		6
36	Seen and Unseen Emotional Style Transfer for Voice Conversion with A New Emotional Speech Dataset. , 2021, , .		53

#	Article	IF	Citations
37	Multi-Target DoA Estimation with an Audio-Visual Fusion Mechanism. , 2021, , .		20
38	Learning Disentangled Feature Representations for Speech Enhancement Via Adversarial Training. , 2021, , .		5
39	Rethinking Benchmarks for Neuromorphic Learning Algorithms. , 2021, , .		0
40	FastTalker: A neural text-to-speech architecture with shallow and group autoregression. Neural Networks, 2021, 141, 306-314.	5.9	8
41	An adaptive transmission line cochlear model based front-end for replay attack detection. Speech Communication, 2021, 132, 114-122.	2.8	1
42	NHSS: A speech and singing parallel database. Speech Communication, 2021, 133, 9-22.	2.8	8
43	HuRAI: A brain-inspired computational model for human-robot auditory interface. Neurocomputing, 2021, 465, 103-113.	5.9	4
44	Transfer Learning From Speech Synthesis to Voice Conversion With Non-Parallel Training Data. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 1290-1302.	5.8	30
45	Expressive TTS Training With Frame and Style Reconstruction Loss. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 1806-1818.	5.8	28
46	Target Speaker Verification With Selective Auditory Attention for Single and Multi-Talker Speech. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 2696-2709.	5.8	11
47	GCC-PHAT with Speech-oriented Attention for Robotic Sound Source Localization. , 2021, , .		7
48	Deep AM-FM: Toolkit for Automatic Dialogue Evaluation. Lecture Notes in Electrical Engineering, 2021, , 53-69.	0.4	8
49	Language Agnostic Speaker Embedding for Cross-Lingual Personalized Speech Generation. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 3427-3439.	5.8	9
50	SLoClas: A Database for Joint Sound Localization and Classification. , 2021, , .		0
51	Low-Latency Auditory Spatial Attention Detection Based on Spectro-Spatial Features from EEG. , 2021, 2021, 5812-5815.		5
52	Expressive Voice Conversion: A Joint Framework for Speaker Identity and Emotional Style Transfer. , 2021, , .		5
53	PL-EESR: Perceptual Loss Based End-to-End Robust Speaker Representation Extraction. , 2021, , .		4
54	Target Language Extraction at Multilingual Cocktail Parties. , 2021, , .		4

#	Article	IF	Citations
55	DEEPA: A Deep Neural Analyzer for Speech and Singing Vocoding. , 2021, , .		0
56	Exploring Teacher-Student Learning Approach for Multi-Lingual Speech-to-Intent Classification. , 2021, , .		0
57	Automatic Leaderboard: Evaluation of Singing Quality Without a Standard Reference. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 13-26.	5.8	7
58	Significance of Subband Features for Synthetic Speech Detection. IEEE Transactions on Information Forensics and Security, 2020, 15, 2160-2170.	6.9	35
59	Multi-Task WaveRNN With an Integrated Architecture for Cross-Lingual Voice Conversion. IEEE Signal Processing Letters, 2020, 27, 1310-1314.	3.6	5
60	On the Importance of Vocal Tract Constriction for Speaker Characterization: The Whispered Speech Study. , 2020, , .		1
61	Modeling Prosodic Phrasing With Multi-Task Learning in Tacotron-Based TTS. IEEE Signal Processing Letters, 2020, 27, 1470-1474.	3.6	10
62	SpEx: Multi-Scale Time Domain Speaker Extraction Network. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 1370-1384.	5.8	66
63	An analysis of speaker dependent models in replay detection. APSIPA Transactions on Signal and Information Processing, 2020, 9, .	3.3	0
64	Teacher-Student Training For Robust Tacotron-Based TTS. , 2020, , .		25
65	Effective Wavenet Adaptation for Voice Conversion with Limited Data. , 2020, , .		4
66	Deep Spiking Neural Networks for Large Vocabulary Automatic Speech Recognition. Frontiers in Neuroscience, 2020, 14, 199.	2.8	59
67	Time-Domain Neural Network Approach for Speech Bandwidth Extension. , 2020, , .		8
68	DeepConversion: Voice conversion with limited parallel training data. Speech Communication, 2020, 122, 31-43.	2.8	14
69	An Efficient and Perceptually Motivated Auditory Neural Encoding and Decoding Algorithm for Spiking Neural Networks. Frontiers in Neuroscience, 2020, 13, 1420.	2.8	21
70	Advances in anti-spoofing: from the perspective of ASVspoof challenges. APSIPA Transactions on Signal and Information Processing, 2020, 9, .	3.3	52
71	Automatic Lyrics Alignment and Transcription in Polyphonic Music: Does Background Music Help?. , 2020, , .		19
72	Assessing the Scope of Generalized Countermeasures for Anti-Spoofing. , 2020, , .		18

#	Article	IF	Citations
73	End-to-End Code-Switching TTS with Cross-Lingual Language Model. , 2020, , .		9
74	An Efficient Threshold-Driven Aggregate-Label Learning Algorithm for Multimodal Information Processing. IEEE Journal on Selected Topics in Signal Processing, 2020, 14, 592-602.	10.8	22
75	Supervised learning in spiking neural networks with synaptic delay-weight plasticity. Neurocomputing, 2020, 409, 103-118.	5.9	36
76	A Cost-Sensitive Deep Belief Network for Imbalanced Classification. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30, 109-122.	11.3	166
77	Neural Population Coding for Effective Temporal Classification. , 2019, , .		22
78	Deep Spiking Neural Network with Spike Count based Learning Rule. , 2019, , .		25
79	Automatic Lyrics-to-audio Alignment on Polyphonic Music Using Singing-adapted Acoustic Models. , 2019, , .		14
80	Cross-lingual Voice Conversion with Bilingual Phonetic Posteriorgram and Average Modeling. , 2019, , .		39
81	Group Sparse Representation With WaveNet Vocoder Adaptation for Spectrum and Prosody Conversion. IEEE/ACM Transactions on Audio Speech and Language Processing, 2019, 27, 1085-1097.	5.8	30
82	Auditory Inspired Spatial Differentiation for Replay Spoofing Attack Detection. , 2019, , .		2
83	Optimization of Speaker Extraction Neural Network with Magnitude and Temporal Spectrum Approximation Loss. , 2019, , .		27
84	RSL2019: A Realistic Speech Localization Corpus. , 2019, , .		5
85	On the Study of Generative Adversarial Networks for Cross-Lingual Voice Conversion. , 2019, , .		14
86	WaveNet Factorization with Singular Value Decomposition for Voice Conversion. , 2019, , .		3
87	SINGAN: Singing Voice Conversion with Generative Adversarial Networks. , 2019, , .		23
88	End-to-End Code-Switching ASR for Low-Resourced Language Pairs. , 2019, , .		13
89	Long Range Acoustic and Deep Features Perspective on ASVspoof 2019. , 2019, , .		36
90	A Modularized Neural Network with Language-Specific Output Layers for Cross-Lingual Voice Conversion. , 2019, , .		9

#	Article	IF	Citations
91	Domain Adversarial Training for Speech Enhancement. , 2019, , .		9
92	Word and Class Common Space Embedding for Code-switch Language Modelling. , 2019, , .		4
93	Speaker-independent Spectral Mapping for Speech-to-Singing Conversion. , 2019, , .		7
94	Multi-band Spectral Entropy Information for Detection of Replay Attacks. , 2019, , .		2
95	Many-to-many Cross-lingual Voice Conversion with a Jointly Trained Speaker Embedding Network. , 2019, , .		4
96	Time-Domain Speaker Extraction Network. , 2019, , .		24
97	Speaker Clustering with Penalty Distance for Speaker Verification with Multi-Speaker Speech. , 2019, , .		0
98	MPD-AL: An Efficient Membrane Potential Driven Aggregate-Label Learning Algorithm for Spiking Neurons. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33, 1327-1334.	4.9	28
99	Competitive STDP-based Feature Representation Learning for Sound Event Classification. , 2019, , .		3
100	Automatic evaluation of end-to-end dialog systems with adequacy-fluency metrics. Computer Speech and Language, 2019, 55, 200-215.	4.3	12
101	Speech-to-Singing Voice Conversion: The Challenges and Strategies for Improving Vocal Conversion Processes. IEEE Signal Processing Magazine, 2019, 36, 95-102.	5.6	12
102	Spike Timing or Rate? Neurons Learn to Make Decisions for Both Through Threshold-Driven Plasticity. IEEE Transactions on Cybernetics, 2019, 49, 2178-2189.	9.5	44
103	The Science of Harmony: A Psychophysical Basis for Perceptual Tensions and Resolutions in Music. Research, 2019, 2019, 2369041.	5.7	6
104	Generalizing I-Vector Estimation for Rapid Speaker Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2018, 26, 749-759.	5.8	15
105	Generative X-Vectors for Text-Independent Speaker Verification. , 2018, , .		8
106	End-to-End Hierarchical Language Identification System. , 2018, , .		3
107	Instantaneous Phase and Excitation Source Features for Detection of Replay Attacks. , 2018, , .		15
108	Extended Constant-Q Cepstral Coefficients for Detection of Spoofing Attacks. , 2018, , .		25

#	Article	IF	Citations
109	A technical framework for automatic perceptual evaluation of singing quality. APSIPA Transactions on Signal and Information Processing, 2018, 7, .	3.3	6
110	Adaptive Wavenet Vocoder for Residual Compensation in GAN-Based Voice Conversion. , 2018, , .		15
111	A Biologically Plausible Speech Recognition Framework Based on Spiking Neural Networks. , 2018, , .		29
112	Farewell Editorial. IEEE/ACM Transactions on Audio Speech and Language Processing, 2018, 26, 2489-2489.	5.8	0
113	Analysis of Speech and Singing Signals for Temporal Alignment. , 2018, , .		11
114	Compensating Utterance Information in Fixed Phrase Speaker Verification. , 2018, , .		2
115	Use of Claimed Speaker Models for Replay Detection. , 2018, , .		5
116	Automatic Evaluation of Singing Quality without a Reference. , 2018, , .		4
117	A Spiking Neural Network Framework for Robust Sound Classification. Frontiers in Neuroscience, 2018, 12, 836.	2.8	69
118	Unsupervised Domain Adaptation via Domain Adversarial Training for Speaker Recognition. , 2018, , .		73
119	Single Channel Speech Separation with Constrained Utterance Level Permutation Invariant Training Using Grid LSTM. , 2018, , .		41
120	An Event-Based Cochlear Filter Temporal Encoding Scheme for Speech Signals. , 2018, , .		11
121	Visual Speech Emotion Conversion using Deep Learning for 3D Talking Head. , 2018, , .		1
122	Re-ranking spoken term detection with acoustic exemplars of keywords. Speech Communication, 2018, 104, 12-23.	2.8	5
123	On the Importance of Analytic Phase of Speech Signals in Spoken Language Recognition. , 2018, , .		4
124	Second Order Factorized Model Adaptation for Short Duration Language Identification. , 2018, , .		1
125	Using language cluster models in hierarchical language identification. Speech Communication, 2018, 100, 30-40.	2.8	2
126	Error Reduction Network for DBLSTM-based Voice Conversion. , 2018, , .		12

#	Article	IF	Citations
127	Many-to-Many Voice Conversion based on Bottleneck Features with Variational Autoencoder for Non-parallel Training Data. , 2018, , .		5
128	On time-frequency mask estimation for MVDR beamforming with application in robust speech recognition. , 2017, , .		46
129	Front-End for Anti-Spoofing Countermeasures in Speaker Verification: Scattering Spectral Decomposition. IEEE Journal on Selected Topics in Signal Processing, 2017, , 1-1.	10.8	10
130	Pairwise learning using multi-lingual bottleneck features for low-resource query-by-example spoken term detection. , 2017, , .		15
131	Multitask Feature Learning for Low-Resource Query-by-Example Spoken Term Detection. IEEE Journal on Selected Topics in Signal Processing, 2017, 11, 1329-1339.	10.8	12
132	Adaptation of PLDA for multi-source text-independent speaker verification. , 2017, , .		0
133	An Exemplar-Based Approach to Frequency Warping for Voice Conversion. IEEE/ACM Transactions on Audio Speech and Language Processing, 2017, 25, 1863-1876.	5.8	32
134	Modeling Latent Topics and Temporal Distance for Story Segmentation of Broadcast News. IEEE/ACM Transactions on Audio Speech and Language Processing, 2017, 25, 112-123.	5.8	13
135	An integrated framework for multimodal human-robot interaction. , 2017, , .		5
136	Facial emotion recognition. , 2017, , .		27
137	A data-driven prognostics framework for tool remaining useful life estimation in tool condition monitoring. , 2017, , .		15
138	Investigation of fixed-dimensional speech representations for real-time speech emotion recognition system. , 2017, , .		3
139	Improving N-gram language modeling for code-switching speech recognition. , 2017, , .		7
140	Named entity transliteration with sequence-to-sequence neural network. , 2017, , .		5
141	Transformation of prosody in voice conversion. , 2017, , .		9
142	Low-resource spoken keyword search strategies in georgian inspired by distinctive feature theory. , 2017, , .		3
143	Improving air traffic control speech intelligibility by reducing speaking rate effectively. , 2017, , .		1
144	A dual alignment scheme for improved speech-to-singing voice conversion. , 2017, , .		11

#	Article	IF	Citations
145	Sparse representation of phonetic features for voice conversion with and without parallel data. , 2017, , .		17
146	Statistical parametric speech synthesis using generative adversarial networks under a multi-task learning framework. , 2017, , .		19
147	Multilingual bottle-neck feature learning from untranscribed speech. , 2017, , .		18
148	Extracting bottleneck features and word-like pairs from untranscribed speech for feature representation. , 2017, , .		13
149	I2R-NUS submission to oriental language recognition AP16-OL7 challenge. , 2017, , .		0
150	A review of the mandarin-english code-switching corpus: SEAME. , 2017, , .		9
151	On the analysis and evaluation of prosody conversion techniques. , 2017, , .		6
152	Perceptual evaluation of singing quality. , 2017, , .		19
153	Combining multiple kernel models for automatic intelligibility detection of pathological speech. , 2016, , .		4
154	Audio and face video emotion recognition in the wild using deep neural networks and small datasets. , 2016, , .		30
155	I-vector based deep neural network acoustic model adaptation using multilingual language resource. , 2016, , .		1
156	Mobile acoustic Emotion Recognition. , 2016, , .		10
157	Computer-assisted pronunciation training: From pronunciation scoring towards spoken language learning. , 2016, , .		19
158	Exemplar-inspired strategies for low-resource spoken keyword search in Swahili. , 2016, , .		12
159	Beamforming networks using spatial covariance features for far-field speech recognition. , 2016, , .		11
160	Cross-lingual deep neural network based submodular unbiased data selection for low-resource keyword search. , 2016, , .		8
161	Spoofing speech detection using temporal convolutional neural network. , 2016, , .		11
162	How the Brain Formulates Memory: A Spatio-Temporal Model Research Frontier. IEEE Computational Intelligence Magazine, 2016, 11, 56-68.	3.2	55

#	Article	IF	Citations
163	Large-scale characterization of non-native Mandarin Chinese spoken by speakers of European origin: Analysis on iCALL. Speech Communication, 2016, 84, 46-56.	2.8	40
164	On the study of replay and voice conversion attacks to text-dependent speaker verification. Multimedia Tools and Applications, 2016, 75, 5311-5327.	3.9	30
165	Approximate search of audio queries by using DTW with phone time boundary and data augmentation. , 2016, , .		8
166	Keyword search using query expansion for graph-based rescoring of hypothesized detections. , 2016, , .		1
167	Spoofing detection from a feature representation perspective. , 2016, , .		28
168	Content-aware local variability vector for speaker verification with short utterance. , 2016, , .		3
169	Country report â€" Singapore. , 2016, , .		0
170	Multi-channel feature adaptation for robust speech recognition. , 2016, , .		0
171	Neural networks based channel compensation for i-vector speaker verification. , 2016, , .		0
172	Exemplar-based sparse representation of timbre and prosody for voice conversion. , 2016, , .		25
173	An expectation-maximization eigenvector clustering approach to direction of arrival estimation of multiple speech sources. , 2016, , .		3
174	A hierarchical framework for language identification. , 2016, , .		11
175	Speech dereverberation for enhancement and recognition using dynamic features constrained deep neural networks and feature adaptation. Eurasip Journal on Advances in Signal Processing, 2016, 2016, .	1.7	27
176	Total Variability Modeling Using Source-Specific Priors. IEEE/ACM Transactions on Audio Speech and Language Processing, 2016, 24, 504-517.	5.8	21
177	Feature Adaptation Using Linear Spectro-Temporal Transform for Robust Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2016, 24, 1006-1019.	5.8	5
178	Single-channel Dereverberation for Distant-Talking Speech Recognition by Combining Denoising Autoencoder and Temporal Structure Normalization. Journal of Signal Processing Systems, 2016, 82, 151-161.	2.1	7
179	A Spiking Neural Network System for Robust Sequence Recognition. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27, 621-635.	11.3	70
180	Exploration of Local Variability in Text-Independent Speaker Verification. Journal of Signal Processing Systems, 2016, 82, 217-228.	2.1	3

#	Article	IF	Citations
181	Joint Chinese word segmentation and punctuation prediction using deep recurrent neural network for social media data. , 2015, , .		3
182	On statistical machine translation method for lexicon refinement in speech recognition. , 2015, , .		0
183	Mapping frames with DNN-HMM recognizer for non-parallel voice conversion. , 2015, , .		5
184	Source-specific informative prior for i-vector extraction. , 2015, , .		3
185	On the study of very low-resource language keyword search. , 2015, , .		3
186	Non-negative matrix factorization using stable alternating direction method of multipliers for source separation. , 2015, , .		6
187	Distance metric learning for kernel density-based acoustic model under limited training data conditions. , 2015, , .		3
188	Mandarinâ€"English code-switching speech corpus in South-East Asia: SEAME. Language Resources and Evaluation, 2015, 49, 581-600.	2.7	28
189	Performance scoring of singing voice. , 2015, , .		2
190	Multilingual exemplar-based acoustic model for the NIST Open KWS 2015 evaluation. , 2015, , .		0
191	The expression of singing emotion - contradicting the constraints of song. , 2015, , .		1
192	A density peak clustering approach to unsupervised acoustic subword units discovery. , 2015, , .		0
193	Towards improving the performance of Vector Space Model for Chinese Frequently Asked Question Answering. , 2015, , .		0
194	Visual Perception Based Engagement Awareness for Multiparty Humanâ€"Robot Interaction. International Journal of Humanoid Robotics, 2015, 12, 1550019.	1.1	2
195	Inaugural Editorial: Embracing New Opportunities for Growth. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 5-6.	5.8	4
196	Channel adaptation of plda for text-independent speaker verification. , 2015, , .		2
197	Robust speech recognition using beamforming with adaptive microphone gains and multichannel noise reduction. , 2015, , .		7
198	Generalised Hough Transform for Speech Pattern Classification. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, , 1-1.	5.8	1

#	Article	IF	Citations
199	Detecting synthetic speech using long term magnitude and phase information. , 2015, , .		4
200	Adequacyâ€"Fluency Metrics: Evaluating MT in the Continuous Space Model Framework. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 472-482.	5.8	50
201	Popular song summarization using chorus section detection from audio signal. , 2015, , .		4
202	Fundamental frequency modeling using wavelets for emotional voice conversion. , 2015, , .		21
203	Speaker Diarization: An Emerging Research. , 2015, , 229-277.		1
204	Introduction to the Special Section on Continuous Space and Related Methods in Natural Language Processing. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 427-430.	5.8	1
205	Acoustic Segment Modeling with Spectral Clustering Methods. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 264-277.	5.8	47
206	A learning-based approach to direction of arrival estimation in noisy and reverberant environments. , 2015, , .		191
207	Formant excursion in singing synthesis. , 2015, , .		4
208	Tokenizing fundamental frequency variation for Mandarin tone error detection. , 2015, , .		3
209	Decoupling Word-Pair Distance and Co-occurrence Information for Effective Long History Context Language Modeling. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 1221-1232.	5.8	1
210	Combining robust spike coding with spiking neural networks for sound event classification. , 2015, , .		3
211	Low-resource keyword search strategies for tamil. , 2015, , .		29
212	Quasi-Factorial Prior for i-vector Extraction. IEEE Signal Processing Letters, 2015, 22, 2484-2488.	3.6	3
213	Language independent query-by-example spoken term detection using N-best phone sequences and partial matching. , 2015, , .		14
214	Relevance factor of maximum a posteriori adaptation for GMMâ€"NAPâ€"SVM in speaker and language recognition. Computer Speech and Language, 2015, 30, 116-134.	4.3	3
215	Spoofing and countermeasures for speaker verification: A survey. Speech Communication, 2015, 66, 130-153.	2.8	405
216	Exemplar-based voice conversion using joint nonnegative matrix factorization. Multimedia Tools and Applications, 2015, 74, 9943-9958.	3.9	20

#	Article	IF	Citations
217	The Pareto Principle Is Everywhere: Finding Informative Sentences for Opinion Summarization Through Leader Detection. Lecture Notes in Social Networks, 2015, , 165-187.	0.1	4
218	CLARA: A Multifunctional Virtual Agent for Conference Support and Touristic Information. , 2015, , 233-239.		10
219	Towards The Use of Probabilistic Spatial Relation Databases in Business Process Modeling. International Journal of Agricultural and Environmental Information Systems, 2015, 6, 50-62.	2.0	0
220	Octave-dependent Probabilistic Latent Semantic Analysis to Chorus Detection of Popular Song. , 2015, , .		0
221	Local variability vector for text-independent speaker verification. , 2014, , .		5
222	Intelligibility detection of pathological speech using asymmetric sparse kernel partial least squares classifier. , 2014, , .		6
223	Emotional facial expression transfer based on temporal restricted Boltzmann machines. , 2014, , .		22
224	Single-channel dereverberation for distant-talking speech recognition by combining denoising autoencoder and temporal structure normalization. , 2014, , .		2
225	Ensemble Nystr&#x00F6;m method for predicting conflict level from speech. , 2014, , .		3
226	Towards better keyword search performance on Malay broadcast news data. , 2014, , .		2
227	Multi-view features in a DNN-CRF model for improved sentence unit detection on English broadcast news. , 2014, , .		3
228	Voice conversion versus speaker verification: an overview. APSIPA Transactions on Signal and Information Processing, 2014, 3, .	3.3	31
229	A study on replay attack and anti-spoofing for text-dependent speaker verification. , 2014, , .		99
230	Feature compensation using linear combination of speaker and environment dependent correction vectors. , 2014, , .		2
231	Subspace Gaussian mixture model for computer-assisted language learning. , 2014, , .		11
232	A discriminatively trained Hough Transform for frame-level phoneme recognition. , 2014, , .		0
233	Improving language modeling by using distance and co-occurrence information of word-pairs and its application to LVCSR. , 2014, , .		1
234	Strategies for Vietnamese keyword search. , 2014, , .		35

#	Article	IF	Citations
235	Discriminative score normalization for keyword search decision. , 2014, , .		7
236	Modelling the alternative hypothesis for text-dependent speaker verification. , 2014, , .		3
237	Minimum divergence estimation of speaker prior in multi-session PLDA scoring. , 2014, , .		20
238	Screen feedback in human-robot interaction: How to enhance robot expressiveness. , 2014, , .		4
239	Acoustic emotion recognition based on fusion of multiple feature-dependent deep Boltzmann machines. , 2014, , .		14
240	Generalization of temporal filter and linear transformation for robust speech recognition. , 2014, , .		1
241	Imposture classification for text-dependent speaker verification. , 2014, , .		5
242	Real-Time Keypoint Recognition Using Restricted Boltzmann Machine. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25, 2119-2126.	11.3	13
243	Learning optimal features for music transcription. , 2014, , .		0
244	Direction-driven navigation using cognitive map for mobile robots. , 2014, , .		15
245	Wikipedia-based Kernels for dialogue topic tracking. , 2014, , .		4
246	Exemplar-Based Sparse Representation With Residual Compensation for Voice Conversion. IEEE/ACM Transactions on Audio Speech and Language Processing, 2014, 22, 1506-1521.	5.8	130
247	Text-dependent speaker verification: Classifiers, databases and RSR2015. Speech Communication, 2014, 60, 56-77.	2.8	188
248	Cross-Lingual Phone Mapping for Large Vocabulary Speech Recognition of Under-Resourced Languages. IEICE Transactions on Information and Systems, 2014, E97.D, 285-295.	0.7	11
249	System and keyword dependent fusion for spoken term detection. , 2014, , .		6
250	A Comparison of Categorical Attribute Data Clustering Methods. Lecture Notes in Computer Science, 2014, , 53-62.	1.3	7
251	Component Pluggable Dialogue Framework and Its Application to Social Robots. , 2014, , 225-237.		3
252	Efficient Language Model Construction for Spoken Dialog Systems by Inducting Language Resources of Different Languages. , 2014, , 101-110.		0

#	Article	IF	Citations
253	Probabilistic Simulation for Probabilistic Data-Aware Business Processes. Lecture Notes in Computer Science, 2014, , 503-515.	1.3	0
254	Continuous attractors of discrete-time recurrent neural networks. Neural Computing and Applications, 2013, 23, 89-96.	5.6	1
255	Optimization Algorithms and Applications for Speech and Language Processing. IEEE Transactions on Audio Speech and Language Processing, 2013, 21, 2231-2243.	3.2	20
256	Spoken Language Recognition With Prosodic Features. IEEE Transactions on Audio Speech and Language Processing, 2013, 21, 1841-1853.	3.2	19
257	A particle filter compensation approach to robust LVCSR. , 2013, , .		0
258	Screen feedback: How to overcome the expressive limitations of a social robot. , 2013, , .		1
259	A dynamic Gaussian process for voice conversion. , 2013, , .		0
260	Conditional restricted Boltzmann machine for voice conversion. , 2013, , .		51
261	Temporal coding of local spectrogram features for robust sound recognition. , 2013, , .		42
262	Dynamics Analysis of a Population Decoding Model. IEEE Transactions on Neural Networks and Learning Systems, 2013, 24, 498-503.	11.3	5
263	Language diarization for conversational code-switch speech with pronunciation dictionary adaptation. , 2013, , .		8
264	Constrained adaptation of histogram equalization for robust speech recognition. , 2013, , .		0
265	Sparse Classifier Fusion for Speaker Verification. IEEE Transactions on Audio Speech and Language Processing, 2013, 21, 1622-1631.	3.2	40
266	Language diarization for code-switch conversational speech. , 2013, , .		16
267	A Spike-Timing-Based Integrated Model for Pattern Recognition. Neural Computation, 2013, 25, 450-472.	2.2	70
268	Shifted-Delta MLP Features for Spoken Language Recognition. IEEE Signal Processing Letters, 2013, 20, 15-18.	3.6	37
269	Spoken Language Recognition: From Fundamentals to Practice. Proceedings of the IEEE, 2013, 101, 1136-1159.	21.3	217
270	Dynamical properties of continuous attractor neural network with background tuning. Neurocomputing, 2013, 99, 439-447.	5.9	2

#	Article	IF	Citations
271	A-STAR: Toward translating Asian spoken languages. Computer Speech and Language, 2013, 27, 509-527.	4.3	6
272	Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. International Journal of Social Robotics, 2013, 5, 171-191.	4.6	211
273	A study on GMM-SVM with adaptive relevance factor and its comparison with i-vector and JFA for speaker recognition. , 2013, , .		2
274	Rapid Feedforward Computation by Temporal Encoding and Learning With Spiking Neurons. IEEE Transactions on Neural Networks and Learning Systems, 2013, 24, 1539-1552.	11.3	120
275	Graph-based informative-sentence selection for opinion summarization. , 2013, , .		13
276	Speech Information Processing: Theory and Applications [Scanning the Issue]. Proceedings of the IEEE, 2013, 101, 1034-1037.	21.3	7
277	The development and analysis of a Malay broadcasr news corpus. , 2013, , .		1
278	Temporal filter design by minimum KL divergence criterion for robust speech recognition. , 2013, , .		3
279	Phonetically-constrained PLDA modeling for text-dependent speaker verification with multiple short utterances. , 2013, , .		35
280	Using parallel tokenizers with DTW matrix combination for low-resource spoken term detection. , 2013, , .		32
281	Recurrent neural network language modeling for code switching conversational speech. , 2013, , .		53
282	Voice conversion and spoofing attack on speaker verification systems. , 2013, , .		41
283	Broadcast news story segmentation using latent topics on data manifold. , 2013, , .		11
284	Minimal-resource phonetic language models to summarize untranscribed speech. , 2013, , .		3
285	Synthetic speech detection using temporal modulation feature. , 2013, , .		77
286	RGB-D based cognitive map building and navigation. , 2013, , .		14
287	Introduction to the Special Section on Large-Scale Optimization for Audio, Speech, and Language Processing. IEEE Transactions on Audio Speech and Language Processing, 2013, 21, 2229-2230.	3.2	0
288	Building Companionship through Human-Robot Collaboration. Lecture Notes in Computer Science, 2013, , 1-7.	1.3	9

#	Article	IF	Citations
289	Precise-Spike-Driven Synaptic Plasticity: Learning Hetero-Association of Spatiotemporal Spike Patterns. PLoS ONE, 2013, 8, e78318.	2.5	137
290	Data-Centric Probabilistic Process: A PhD. Symposium Paper for ICSOC. Lecture Notes in Computer Science, 2013, , 406-411.	1.3	0
291	Vision-based attention estimation and selection for social robot to perform natural interaction in the open world. , 2012, , .		9
292	Mixture of Factor Analyzers Using Priors From Non-Parallel Speech for Voice Conversion. IEEE Signal Processing Letters, 2012, 19, 914-917.	3.6	71
293	Joint spectral and temporal normalization of features for robust recognition of noisy and reverberated speech. , 2012, , .		6
294	Acoustic TextTiling for story segmentation of spoken documents. , 2012, , .		14
295	An acoustic segment modeling approach to query-by-example spoken term detection. , 2012, , .		44
296	MODULAR IK: A ROBUST INVERSE KINEMATIC ALGORITHM FOR GESTURE IMITATION IN AN UPPER-BODY HUMANOID ROBOT. International Journal of Humanoid Robotics, 2012, 09, 1250010.	1.1	4
297	Lasso environment model combination for robust speech recognition. , 2012, , .		7
298	A Phone Mapping Technique for Acoustic Modeling of Under-Resourced Languages. , 2012, , .		4
299	A bootstrapping approach for SLU portability to a new language by inducting unannotated user queries. , 2012, , .		7
300	Selective Gammatone Envelope Feature for Robust Sound Event Recognition. IEICE Transactions on Information and Systems, 2012, E95.D, 1229-1237.	0.7	1
301	Vulnerability of speaker verification systems against voice conversion spoofing attacks: The case of telephone speech. , 2012, , .		119
302	Adaptive control for robot manipulators under ellipsoidal task space constraints. , 2012, , .		4
303	A first speech recognition system for Mandarin-English code-switch conversational speech. , 2012, , .		81
304	An analysis of vector Taylor series model compensation for non-stationary noise in speech recognition. , 2012, , .		1
305	Phonotactic spoken language recognition: Using diversely adapted acoustic models in parallel phone recognizers. , 2012, , .		0
306	A study of F0 modelling and generation with lyrics and shape characterization for singing voice synthesis. , 2012, , .		2

#	Article	IF	Citations
307	Generalized F0 modelling with absolute and relative pitch features for singing voice synthesis. , 2012, , .		10
308	Collection and annotation of Malay conversational speech corpus. , 2012, , .		2
309	Context dependant phone mapping for cross-lingual acoustic modeling. , 2012, , .		7
310	Robust Multiperson Detection and Tracking for Mobile Service and Social Robots. IEEE Transactions on Systems, Man, and Cybernetics, 2012, 42, 1398-1412.	5.0	28
311	Discriminative feature extraction for speech recognition using continuous output codes. Pattern Recognition Letters, 2012, 33, 1703-1709.	4.2	10
312	I-vectors in the context of phonetically-constrained short utterances for speaker verification. , 2012, , .		34
313	Learning regional transliteration variants. Information Processing and Management, 2012, 48, 154-169.	8.6	1
314	Gesture Recognition Based on Localist Attractor Networks with Application to Robot Control [Application Notes]. IEEE Computational Intelligence Magazine, 2012, 7, 64-74.	3.2	19
315	Speaker Clustering and Cluster Purification Methods for RT07 and RT09 Evaluation Meeting Data. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 461-473.	3.2	11
316	Bitext Dependency Parsing With Auto-Generated Bilingual Treebank. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 1461-1472.	3.2	1
317	Low-Variance Multitaper MFCC Features: A Case Study in Robust Speaker Verification. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 1990-2001.	3.2	106
318	Broadcast News Story Segmentation Using Conditional Random Fields and Multimodal Features. IEICE Transactions on Information and Systems, 2012, E95.D, 1206-1215.	0.7	13
319	Spectrogram Image Feature for Sound Event Classification in Mismatched Conditions. IEEE Signal Processing Letters, 2011, 18, 130-133.	3.6	166
320	Language Identification: A Tutorial. IEEE Circuits and Systems Magazine, 2011, 11, 82-108.	2.3	104
321	Error Corrective Fusion of Classifier Scores for Spoken Language Recognition. IEICE Transactions on Information and Systems, 2011, E94-D, 2503-2512.	0.7	0
322	Speaker Verification With Feature-Space MAPLR Parameters. IEEE Transactions on Audio Speech and Language Processing, 2011, 19, 505-515.	3.2	7
323	Using Discrete Probabilities With Bhattacharyya Measure for SVM-Based Speaker Verification. IEEE Transactions on Audio Speech and Language Processing, 2011, 19, 861-870.	3.2	12
324	Sound Event Recognition With Probabilistic Distance SVMs. IEEE Transactions on Audio Speech and Language Processing, 2011, 19, 1556-1568.	3.2	49

#	Article	IF	Citations
325	A Maximum-Entropy Segmentation Model for Statistical Machine Translation. IEEE Transactions on Audio Speech and Language Processing, 2011, 19, 2494-2505.	3.2	11
326	Towards an Effective Design of Social Robots. International Journal of Social Robotics, 2011, 3, 333-335.	4.6	48
327	Online learning and stimulus-driven responses of neurons in visual cortex. Cognitive Neurodynamics, 2011, 5, 77-85.	4.0	1
328	A cross-domain adaptation method for sentiment classification using probabilistic latent analysis. , 2011, , .		44
329	Jump Function Kolmogorov for overlapping audio event classification. , 2011, , .		7
330	Factored covariance modeling for text-independent speaker verification. , 2011, , .		0
331	Maximum likelihood adaptation of histogram equalization with constraint for robust speech recognition. , 2011, , .		11
332	Classifier subset selection and fusion for speaker verification. , 2011, , .		3
333	Score fusion and calibration in multiple language detectors with large performance variation. , 2011, , .		0
334	Beat space segmentation and octave scale cepstral feature for sung language recognition in pop music. ACM Transactions on Multimedia Computing, Communications and Applications, 2011, 7, 1-19.	4.3	0
335	Probabilistic distance SVM with Hellinger-Exponential Kernel for sound event classification. , 2011, , .		5
336	Effective Large Scale Text Retrieval via Learning Risk-Minimization and Dependency-Embedded Model. Lecture Notes in Computer Science, 2011, , 99-110.	1.3	0
337	TEXT-INDEPENDENT SPEAKER RECOGNITION. , 2011, , 3-29.		2
338	An overview of text-independent speaker recognition: From features to supervectors. Speech Communication, 2010, 52, 12-40.	2.8	1,149
339	Adaptive learning tracking control of robotic manipulators with uncertainties. Journal of Control Theory and Applications, 2010, 8, 160-165.	0.8	8
340	A Study on the Generalization Capability of Acoustic Models for Robust Speech Recognition. IEEE Transactions on Audio Speech and Language Processing, 2010, 18, 1158-1169.	3.2	20
341	GMM-SVM Kernel With a Bhattacharyya-Based Distance for Speaker Recognition. IEEE Transactions on Audio Speech and Language Processing, 2010, 18, 1300-1312.	3.2	64
342	Linguistically Annotated Reordering: Evaluation and Analysis. Computational Linguistics, 2010, 36, 535-568.	3.3	3

#	Article	IF	Citations
343	Prosodic attribute model for spoken language identification. , 2010, , .		13
344	Error corrective classifier fusion for spoken Language Recognition. , 2010, , .		0
345	Scream detection for home applications. , 2010, , .		25
346	Speaker diarization system for RT07 and RT09 meeting room audio. , 2010, , .		19
347	Framewise Phone Classification Using Weighted Fuzzy Classification Rules. , 2010, , .		0
348	Soft margin estimation of Gaussian mixture model parameters for spoken language recognition. , 2010, , .		1
349	Nonlinear Control of a Robot Manipulator with Time-Varying Uncertainties. Lecture Notes in Computer Science, 2010, , 202-211.	1.3	1
350	Tuning phone decoders for language identification. , 2010, , .		2
351	Voice conversion: From spoken vowels to singing vowels. , 2010, , .		16
352	An acoustic segment model approach to incorporating temporal information into speaker modeling for text-independent speaker recognition. , 2010, , .		10
353	Semi-supervised learning of language model using unsupervised topic model. , 2010, , .		0
354	Word level automatic alignment of music and lyrics using vocal synthesis. ACM Transactions on Multimedia Computing, Communications and Applications, 2010, 6, 1-16.	4.3	2
355	Memory Dynamics in Attractor Networks with Saliency Weights. Neural Computation, 2010, 22, 1899-1926.	2.2	26
356	How humans behave and evaluate a social robot in real-environment settings. , 2010, , .		1
357	Statistical lattice-based spoken document retrieval. ACM Transactions on Information Systems, 2010, 28, 1-30.	4.9	25
358	TechWare: Speaker and Spoken Language Recognition Resources [Best of the Web. IEEE Signal Processing Magazine, 2010, , .	5.6	3
359	Generating emotional speech from neutral speech. , 2010, , .		8
360	Adaptive admittance control of a robot manipulator under task space constraint. , 2010, , .		15

#	Article	IF	Citations
361	Frame selection of interview channel for NIST speaker recognition evaluation. , 2010, , .		3
362	A Discrete-Time Neural Network for Optimization Problems With Hybrid Constraints. IEEE Transactions on Neural Networks, 2010, 21, 1184-1189.	4.2	13
363	Feature integration for heart sound biometrics. , 2010, , .		2
364	The psychoacoustic approach towards enhancing speech intelligibility in noise. , 2010, , .		0
365	Aligning singing voice with MIDI melody using synthesized audio signal. , 2010, , .		2
366	Building topic mixture language models using the document soft classification notion of topic models. , 2010, , .		1
367	UBM data selection for effective speaker modeling. , 2010, , .		2
368	The I4U system in NIST 2008 speaker recognition evaluation. , 2009, , .		18
369	Cross-validation of multiple language recognition systems using pseudo keys. , 2009, , .		0
370	Analysis and Selection of Prosodic Features for Language Identification. , 2009, , .		13
371	Acoustic segment modeling for speaker recognition. , 2009, , .		8
372	Subspace construction and selection for speaker recognition. , 2009, , .		0
373	Control of nonlinear systems with time-varying output constraints. , 2009, , .		29
374	Refining Unit Boundaries for Mandarin Text-to-Speech Database. , 2009, , .		0
375	Cluster criterion functions in spectral subspace and their application in speaker clustering. , 2009, , .		3
376	MASS: A Malay language LVCSR corpus resource. , 2009, , .		25
377	Joint map adaptation of feature transformation and Gaussian Mixture Model for speaker recognition. , 2009, , .		8
378	Efficient sparse self-similarity matrix construction for repeating sequence detection. , 2009, , .		1

#	Article	IF	Citations
379	Speaker diarization in meeting audio. , 2009, , .		9
380	Nonlinear control approaches for SI engine model with uncertainties. , 2009, , .		3
381	Evaluation of a fused FM and cepstral-based speaker recognition system on the NIST 2008 SRE. , 2009, , .		2
382	Adaptive tracking control of uncertain robotic manipulators in a constrained task space. , 2009, , .		1
383	A Lattice-Based Phonotactic Language Recognition System with CMLLR Adaptation and Its Implementation Issues. , 2009, , .		3
384	Exploiting prosodic information for Speaker Recognition. , 2009, , .		1
385	A GMM supervector Kernel with the Bhattacharyya distance for SVM based speaker recognition. , 2009, , .		10
386	Introduction to the Special Issue on Recent Advances in Asian Language Spoken Document Retrieval. ACM Transactions on Asian Language Information Processing, 2009, 8, 1-3.	0.8	1
387	A Target-Oriented Phonotactic Front-End for Spoken Language Recognition. IEEE Transactions on Audio Speech and Language Processing, 2009, 17, 1335-1347.	3.2	16
388	Jump Function Kolmogorov for Audio Classification in Noise-Mismatch Conditions. IEEE Transactions on Signal Processing, 2009, 57, 2908-2918.	5.3	1
389	Semi-supervised Learning of Domain-Specific Language Models from General Domain Data. , 2009, , .		0
390	The Asian network-based speech-to-speech translation system. , 2009, , .		8
391	An SVM Kernel With GMM-Supervector Based on the Bhattacharyya Distance for Speaker Recognition. IEEE Signal Processing Letters, 2009, 16, 49-52.	3.6	68
392	A study on hidden Markov model's generalization capability for speech recognition. , 2009, , .		2
393	A life-size robotic lion dance system with integrated motion control. , 2009, , .		2
394	Harvesting Regional Transliteration Variants with Guided Search. Lecture Notes in Computer Science, 2009, , 133-144.	1.3	2
395	An Interactive Robot Butler. Lecture Notes in Computer Science, 2009, , 385-394.	1.3	3
396	Experiences with a Barista Robot, FusionBot. Communications in Computer and Information Science, 2009, , 140-151.	0.5	8

#	Article	IF	Citations
397	A Survey of Music Structure Analysis Techniques for Music Applications. Studies in Computational Intelligence, 2009, , 551-577.	0.9	1
398	Active learning for constructing transliteration lexicons from the Web. Journal of the Association for Information Science and Technology, 2008, 59, 126-135.	2.6	6
399	Normalization of the Speech Modulation Spectra for Robust Speech Recognition. IEEE Transactions on Audio Speech and Language Processing, 2008, 16, 1662-1674.	3.2	59
400	Optimizing the Performance of Spoken Language Recognition With Discriminative Training. IEEE Transactions on Audio Speech and Language Processing, 2008, 16, 1642-1653.	3.2	7
401	On Acoustic Diversification Front-End for Spoken Language Identification. IEEE Transactions on Audio Speech and Language Processing, 2008, 16, 1029-1037.	3.2	23
402	Vector-Based Spoken Language Classification. Springer Handbooks, 2008, , 825-840.	0.6	0
403	Target-oriented phone tokenizers for spoken language recognition. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	7
404	Self-Organized Clustering for Feature Mapping in Language Recognition. , 2008, , .		2
405	Using Pseudo-Key for Language Recognition System Design. , 2008, , .		0
406	Predicting Spectral and Prosodic Parameters for Unit Selection in Speech Synthesis. , 2008, , .		0
407	PLSA Based Topic Mixture Language Modeling Approach. , 2008, , .		0
408	An Efficient Feature Selection Method for Speaker Recognition. , 2008, , .		11
409	Discriminative Output Coding Features for Speech Recognition. , 2008, , .		1
410	Fuzzy rule selection using Iterative Rule Learning for speech data classification. , 2008, , .		2
411	Effect of Feature Smoothing for Robust Speech Recognition. , 2008, , .		0
412	Classification of speech transmission channels: Landline, GSM and VoIP networks. , 2008, , .		0
413	Spoken Language recognition using support vector machines with generative front-end. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	3
414	On fusion of timbre-motivated features for singing voice detection and singer identification. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	16

#	Article	IF	Citations
415	Robust phone set mapping using decision tree clustering for cross-lingual phone recognition. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	9
416	Singing voice detection in pop songs using co-training algorithm. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	4
417	Mining Live Transliterations Using Incremental Learning Algorithms. International Journal of Computer Processing of Languages, 2008, 21, 183-203.	0.3	3
418	A lattice-based approach to query-by-example spoken document retrieval. , 2008, , .		37
419	Speech enhancement for telephony name speech recognition. , 2008, , .		0
420	Discriminative learning for optimizing detection performance in spoken language recognition. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	1
421	Unsupervised pronunciation grammar growing using knowledge-based and data-driven approaches. , 2008, , .		1
422	Proportionate subband adaptive filtering. , 2008, , .		2
423	Exploring Perceptual Based Timbre Feature for Singer Identification. Lecture Notes in Computer Science, 2008, , 159-171.	1.3	7
424	Bicluster Analysis of Currency Exchange Rates. , 2008, , 19-34.		5
425	A phonetic similarity model for automatic extraction of transliteration pairs. ACM Transactions on Asian Language Information Processing, 2007, 6, 6.	0.8	18
426	Singing voice detection using perceptually-motivated features. , 2007, , .		1
427	A Vector-Based Approach to Broadcast Audio Database Indexing and Retrieval. , 2007, , .		0
428	A Generalized Feature Transformation Approach for Channel Robust Speaker Verification. , 2007, , .		8
429	<title>Speaker diarization system on the 2007 NIST rich transcription meeting recognition evaluation</title>. , 2007, , .		0
430	Temporal Structure Normalization of Speech Feature for Robust Speech Recognition. IEEE Signal Processing Letters, 2007, 14, 500-503.	3.6	30
431	Spoken Language Recognition with Relevance Feedback. , 2007, , .		2
432	Normalizing the Speech Modulation Spectrum for Robust Speech Recognition. , 2007, , .		11

#	Article	IF	Citations
433	Discriminative Vector for Spoken Language Recognition. , 2007, , .		7
434	Exploring Vibrato-Motivated Acoustic Features for Singer Identification. IEEE Transactions on Audio Speech and Language Processing, 2007, 15, 519-530.	3.2	56
435	A Vector Space Modeling Approach to Spoken Language Identification. IEEE Transactions on Audio Speech and Language Processing, 2007, 15, 271-284.	3.2	166
436	Spoken Language Recognition Using Ensemble Classifiers. IEEE Transactions on Audio Speech and Language Processing, 2007, 15, 2053-2062.	3.2	23
437	Speaker Diarization Using Direction of Arrival Estimate and Acoustic Feature Information: The I2R-NTU Submission for the NIST RT 2007 Evaluation. Lecture Notes in Computer Science, 2007, , 484-496.	1.3	6
438	A Hierarchical Approach for Music Chord Modeling Based on the Analysis of Tonal Characteristics. , 2006, , .		1
439	Syllabic level automatic synchronization of music signals and text lyrics. , 2006, , .		16
440	Music structure based vector space retrieval. , 2006, , .		21
441	Language Recognition Based on Score Distribution Feature Vectors and Discriminative Classifier Fusion. , 2006, , .		7
442	Fusion of Acoustic and Tokenization Features for Speaker Recognition. Lecture Notes in Computer Science, 2006, , 566-577.	1.3	6
443	The IIR Submission to CSLP 2006 Speaker Recognition Evaluation. Lecture Notes in Computer Science, 2006, , 494-505.	1.3	0
444	Vector Autoregressive Model for Missing Feature Reconstruction. Lecture Notes in Computer Science, 2006, , 315-324.	1.3	0
445	TRANSLITERATION. , 2006, , 341-364.		0
446	A phonotactic-semantic paradigm for automatic spoken document classification. , 2005, , .		10
447	Phrase-Based Statistical Machine Translation: A Level of Detail Approach. Lecture Notes in Computer Science, 2005, , 576-587.	1.3	4
448	A phonotactic language model for spoken language identification. , 2005, , .		26
449	An objective measure for assessment of a corpus-based text-to-speech system. , 0, , .		1
450	Using Local & Global Phonotactic Features in Chinese Dialect Identification. , 0, , .		2

#	Article	IF	Citations
451	Broadcast News Segmentation By Audio Type Analysis. , 0, , .		6
452	Integrating Acoustic, Prosodic and Phonotactic Features for Spoken Language Identification. , 0, , .		46
453	Neural Speaker Extraction with Speaker-Speech Cross-Attention Network. , 0, , .		6
454	Universal Speaker Extraction in the Presence and Absence of Target Speakers for Speech of One and Two Talkers. , 0, , .		11
455	The reddots data collection for speaker recognition. , 0, , .		47
456	Deep Bidirectional LSTM Modeling of Timbre and Prosody for Emotional Voice Conversion. , 0, , .		39
457	Semi-Supervised and Cross-Lingual Knowledge Transfer Learnings for DNN Hybrid Acoustic Models Under Low-Resource Conditions. , 0, , .		10
458	SingaKids-Mandarin: Speech Corpus of Singaporean Children Speaking Mandarin Chinese. , 0, , .		5
459	Context Aware Mispronunciation Detection for Mandarin Pronunciation Training. , 0, , .		3
460	Unsupervised Bottleneck Features for Low-Resource Query-by-Example Spoken Term Detection. , 0, , .		29
461	Learning Neural Network Representations Using Cross-Lingual Bottleneck Features with Word-Pair Information. , 0, , .		16
462	Out of Set Language Modelling in Hierarchical Language Identification. , 0, , .		4
463	The 2015 NIST Language Recognition Evaluation: The Shared View of I2R, Fantastic4 and SingaMS. , 0, , .		16
464	Rescoring Hypothesized Detections of Out-of-Vocabulary Keywords Using Subword Samples. , 0, , .		2
465	Toward High-Performance Language-Independent Query-by-Example Spoken Term Detection for MediaEval 2015: Post-Evaluation Analysis. , 0, , .		13
466	An Investigation of Spoofing Speech Detection Under Additive Noise and Reverberant Conditions. , 0, , .		12
467	A DNN-HMM Approach to Story Segmentation. , 0, , .		17
468	Weighted Spatial Covariance Matrix Estimation for MUSIC Based TDOA Estimation of Speech Source. , 0, , .		12

#	Article	IF	Citations
469	Learning Acoustic Word Embeddings with Temporal Context for Query-by-Example Speech Search. , 0, , .		16
470	A Voice Conversion Framework with Tandem Feature Sparse Representation and Speaker-Adapted WaveNet Vocoder. , 0, , .		28
471	Wavelet Analysis of Speaker Dependent and Independent Prosody for Voice Conversion. , 0, , .		11
472	Code-Switching Detection Using ASR-Generated Language Posteriors. , 0, , .		1
473	Joint Training Framework for Text-to-Speech and Voice Conversion Using Multi-Source Tacotron and WaveNet. , 0, , .		32
474	Linguistically Motivated Parallel Data Augmentation for Code-Switch Language Modeling. , 0, , .		15
475	On the End-to-End Solution to Mandarin-English Code-Switching Speech Recognition. , 0, , .		28
476	Long Range Acoustic Features for Spoofed Speech Detection. , 0, , .		30
477	On the Importance of Audio-Source Separation for Singer Identification in Polyphonic Music. , 0, , .		13
478	Multi-Level Adaptive Speech Activity Detector for Speech in Naturalistic Environments. , 0, , .		7
479	A Combination of Model-Based and Feature-Based Strategy for Speech-to-Singing Alignment. , 0, , .		2
480	An Adaptive-Q Cochlear Model for Replay Spoofing Detection. , 0, , .		1
481	SpEx+: A Complete Time Domain Speaker Extraction Network. , 0, , .		53
482	Audio-Visual Speaker Recognition with a Cross-Modal Discriminative Network. , 0, , .		19
483	Converting Anyone's Emotion: Towards Speaker-Independent Emotional Voice Conversion. , 0, , .		13
484	Low Latency Auditory Attention Detection with Common Spatial Pattern Analysis of EEG Signals. , 0, , .		12
485	I2R Submission to the 2015 NIST Language Recognition I-vector Challenge. , 0, , .		3
486	Rapid Computation of I-vector. , 0, , .		6

#	Article	IF	Citations
487	Transforming Spectrum and Prosody for Emotional Voice Conversion with Non-Parallel Training Data. , 0, , .		31
488	WaveTTS: Tacotron-based TTS with Joint Time-Frequency Domain Loss. , 0, , .		10
489	Speaker cluster based GMM tokenization for speaker recognition. , 0, , .		10
490	Vector-based spoken language recognition using output coding. , 0, , .		4
491	A GMM-based probabilistic sequence kernel for speaker verification. , 0, , .		11
492	Using MAP estimation of feature transformation for speaker recognition. , 0, , .		6
493	Characterizing speech utterances for speaker verification with sequence kernel SVM. , 0, , .		9
494	T-test distance and clustering criterion for speaker diarization. , 0, , .		23
495	Context-sensitive probabilistic phone mapping model for cross-lingual speech recognition. , 0, , .		12
496	Unit selection based speech synthesis for poor channel condition. , 0, , .		1
497	Target-aware language models for spoken language recognition. , 0, , .		4
498	Text-independent F0 transformation with non-parallel data for voice conversion. , 0, , .		16
499	Speaker characterization using long-term and temporal information. , 0, , .		7
500	SEAME: a Mandarin-English code-switching speech corpus in south-east asia. , 0, , .		35
501	Towards long-range prosodic attribute modeling for language recognition. , 0, , .		2
502	Selecting phonotactic features for language recognition. , 0, , .		8
503	PLDA modeling in i-vector and supervector space for speaker verification. , 0, , .		27
504	RSR2015: database for text-dependent speaker verification using multiple pass-phrases. , 0, , .		46

#	Article	IF	Citations
505	Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition. , 0, , .		86
506	Vulnerability evaluation of speaker verification under voice conversion spoofing: the effect of text constraints. , 0, , .		15
507	Unsupervised mining of acoustic subword units with segment-level Gaussian posteriorgrams. , 0, , .		14
508	A graph-based Gaussian component clustering approach to unsupervised acoustic modeling. , 0, , .		7
509	Rapid Update of Multilingual Deep Neural Network for Low-Resource Keyword Search. , 0, , .		5
510	SERAPHIM: A Wavetable Synthesis System with 3D Lip Animation for Real-Time Speech and Singing Applications on Mobile Platforms. , 0, , .		1
511	Denoising Recurrent Neural Network for Deep Bidirectional LSTM Based Voice Conversion. , 0, , .		5
512	Multimodal Prediction of Affective Dimensions via Fusing Multiple Regression Techniques. , 0, , .		1
513	Investigating Scalability in Hierarchical Language Identification System. , 0, , .		1
514	Gain Compensation for Fast i-Vector Extraction Over Short Duration. , 0, , .		2
515	Average Modeling Approach to Voice Conversion with Non-Parallel Data. , 0, , .		20
516	A Shifted Delta Coefficient Objective for Monaural Speech Separation Using Multi-task Learning. , 0, , .		8
517	Automatic Pronunciation Evaluation of Singing. , 0, , .		9
518	Co-whitening of I-vectors for Short and Long Duration Speaker Verification. , 0, , .		1
519	Target Speaker Extraction for Multi-Talker Speaker Verification. , 0, , .		12
520	Large-Scale Speaker Diarization of Radio Broadcast Archives. , 0, , .		0
521	A Speaker-Dependent WaveNet for Voice Conversion with Non-Parallel Data. , 0, , .		16
522	A Unified Framework for Speaker and Utterance Verification. , 0, , .		4

#	Article	IF	Citations
523	Acoustic Modeling for Automatic Lyrics-to-Audio Alignment. , 0, , .		10
524	Multi-Graph Decoding for Code-Switching ASR. , 0, , .		3
525	VQVAE Unsupervised Unit Discovery and Multi-Scale Code2Spec Inverter for Zerospeech Challenge 2019. , 0, , .		26
526	Instantaneous Phase and Long-Term Acoustic Cues for Orca Activity Detection. , 0, , .		3
527	Speaker-Utterance Dual Attention for Speaker and Utterance Verification. , 0, , .		6
528	Multi-Modal Attention for Speech Emotion Recognition. , 0, , .		35