

# Hsin-Min Wang

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/8904087/publications.pdf>

Version: 2024-02-01

212  
papers

2,650  
citations

377584

21  
h-index

425179

34  
g-index

213  
all docs

213  
docs citations

213  
times ranked

1509  
citing authors

#	ARTICLE	IF	CITATIONS
1	Improved Lite Audio-Visual Speech Enhancement. IEEE/ACM Transactions on Audio Speech and Language Processing, 2022, 30, 1345-1359.	4.0	12
2	SVSNet: An End-to-End Speaker Voice Similarity Assessment Model. IEEE Signal Processing Letters, 2022, 29, 767-771.	2.1	1
3	Partially Fake Audio Detection by Self-Attention-Based Fake Span Discovery. , 2022, , .		8
4	EMGSE: Acoustic/EMG Fusion for Multimodal Speech Enhancement. , 2022, , .		0
5	Learning to Visualize Music Through Shot Sequence for Automatic Concert Video Mashup. IEEE Transactions on Multimedia, 2021, 23, 1731-1743.	5.2	3
6	Improving the Intelligibility of Speech for Simulated Electric and Acoustic Stimulation Using Fully Convolutional Neural Networks. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29, 184-195.	2.7	12
7	MoEVC: A Mixture of Experts Voice Conversion System With Sparse Gating Mechanism for Online Computation Acceleration. , 2021, , .		1
8	Melody Harmonization Using Orderless Nade, Chord Balancing, and Blocked Gibbs Sampling. , 2021, , .		5
9	Speech Enhancement with Zero-Shot Model Selection. , 2021, , .		3
10	Investigation of a Single-Channel Frequency-Domain Speech Enhancement Network to Improve End-to-End Bengali Automatic Speech Recognition Under Unseen Noisy Conditions. , 2021, , .		1
11	Mandarin Electrolaryngeal Speech Voice Conversion with Sequence-to-Sequence Modeling. , 2021, , .		3
12	HASA-Net: A Non-Intrusive Hearing-Aid Speech Assessment Network. , 2021, , .		3
13	Speech Enhancement Based on Denoising Autoencoder With Multi-Branched Encoders. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 2756-2769.	4.0	17
14	WaveCRN: An Efficient Convolutional Recurrent Neural Network for End-to-End Speech Enhancement. IEEE Signal Processing Letters, 2020, 27, 2149-2153.	2.1	45
15	Self-Supervised Denoising Autoencoder with Linear Regression Decoder for Speech Enhancement. , 2020, , .		16
16	Statistics Pooling Time Delay Neural Network Based on X-Vector for Speaker Verification. , 2020, , .		9
17	Combining Deep Embeddings of Acoustic and Articulatory Features for Speaker Identification. , 2020, , .		6
18	Multichannel Speech Enhancement by Raw Waveform-Mapping Using Fully Convolutional Networks. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 1888-1900.	4.0	27

#	ARTICLE	IF	CITATIONS
19	Unsupervised Representation Disentanglement Using Cross Domain Features and Adversarial Learning in Variational Autoencoder Based Voice Conversion. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020, 4, 468-479.	3.4	22
20	ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech. Computer Speech and Language, 2020, 64, 101114.	2.9	130
21	Subspace-Based Representation and Learning for Phonotactic Spoken Language Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 3065-3079.	4.0	8
22	Using Taigi Dramas with Mandarin Chinese Subtitles to Improve Taigi Speech Recognition. , 2020, , .		1
23	Reinforcement Learning Based Speech Enhancement for Robust Speech Recognition. , 2019, , .		17
24	Oriental COCOSDA - Country report 2019 language resources developed in Taiwan. , 2019, , .		0
25	Audio-Visual Speech Enhancement using Hierarchical Extreme Learning Machine. , 2019, , .		1
26	Spoken Multiple-Choice Question Answering Using Multimodal Convolutional Neural Networks. , 2019, , .		2
27	Compressed Multimodal Hierarchical Extreme Learning Machine for Speech Enhancement. , 2019, , .		0
28	Sequential Speaker Embedding and Transfer Learning for Text-Independent Speaker Identification. , 2019, , .		0
29	Investigation of Neural Network Approaches for Unified Spectral and Prosodic Feature Enhancement. , 2019, , .		1
30	Multi-task Learning for Acoustic Modeling Using Articulatory Attributes. , 2019, , .		2
31	Improving Automatic Jazz Melody Generation by Transfer Learning Techniques. , 2019, , .		8
32	Refined WaveNet Vocoder for Variational Autoencoder Based Voice Conversion. , 2019, , .		9
33	Coherent Deep-Net Fusion To Classify Shots In Concert Videos. IEEE Transactions on Multimedia, 2018, , 1-1.	5.2	13
34	An Information Distillation Framework for Extractive Summarization. IEEE/ACM Transactions on Audio Speech and Language Processing, 2018, 26, 161-170.	4.0	26
35	Audio-Visual Speech Enhancement Using Multimodal Deep Convolutional Neural Networks. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, 2, 117-128.	3.4	151
36	Voice Conversion Based on Cross-Domain Features Using Variational Auto Encoders. , 2018, , .		24

#	ARTICLE	IF	CITATIONS
37	Seethevoice: Learning from Music to Visual Storytelling of Shots. , 2018, , .		4
38	Deep Denoising Autoencoder Based Post Filtering for Speech Enhancement. , 2018, , .		5
39	Automatic Detection of Speech Under Cold Using Discriminative Autoencoders and Strength Modeling with Multiple Sub-Dictionary Generation. , 2018, , .		2
40	Essence Vector-Based Query Modeling for Spoken Document Retrieval. , 2018, , .		1
41	Speech emotion recognition with skew-robust neural networks. , 2017, , .		8
42	A Position-Aware Language Modeling Framework for Extractive Broadcast News Speech Summarization. ACM Transactions on Asian and Low-Resource Language Information Processing, 2017, 16, 1-13.	1.3	1
43	Discriminative autoencoders for speaker verification. , 2017, , .		7
44	A locality-preserving essence vector modeling framework for spoken document retrieval. , 2017, , .		5
45	Leveraging manifold learning for extractive broadcast news summarization. , 2017, , .		3
46	Deep-net fusion to classify shots in concert videos. , 2017, , .		4
47	A locally linear embedding based postfiltering approach for speech enhancement. , 2017, , .		1
48	Automatic Music Video Generation Based on Simultaneous Soundtrack Recommendation and Video Editing. , 2017, , .		6
49	Exploring microscopic fluctuation of facial expression for mood disorder classification. , 2017, , .		5
50	A chatbot using LSTM-based multi-layer embedding for elderly care. , 2017, , .		41
51	Neural relevance-aware query modeling for spoken document retrieval. , 2017, , .		3
52	Fast locally linear embedding algorithm for exemplar-based voice conversion. , 2017, , .		0
53	Personality trait perception from speech signals using multiresolution analysis and convolutional neural networks. , 2017, , .		5
54	DEMV-matchmaker: Emotional temporal course representation and deep similarity matching for automatic music video generation. , 2016, , .		11

#	ARTICLE	IF	CITATIONS
55	A novel paragraph embedding method for spoken document summarization. , 2016, , .		0
56	Voice conversion from non-parallel corpora using variational auto-encoder. , 2016, , .		145
57	Novel Word Embedding and Translation-based Language Modeling for Extractive Speech Summarization. , 2016, , .		3
58	Exploiting graph regularized nonnegative matrix factorization for extractive speech summarization. , 2016, , .		0
59	Audio-visual speech enhancement using deep neural networks. , 2016, , .		16
60	Exploring the use of unsupervised query modeling techniques for speech recognition and summarization. Speech Communication, 2016, 80, 49-59.	1.6	9
61	Improved spoken document summarization with coverage modeling techniques. , 2016, , .		3
62	Alignment of Lyrics With Accompanied Singing Audio Based on Acoustic-Phonetic Vowel Likelihood Modeling. IEEE/ACM Transactions on Audio Speech and Language Processing, 2016, 24, 1998-2008.	4.0	9
63	Automatic Music Video Generation Based on Emotion-Oriented Pseudo Song Prediction and Matching. , 2016, , .		7
64	Dictionary update for NMF-based voice conversion using an encoder-decoder network. , 2016, , .		3
65	Affective Music Information Retrieval. Human-computer Interaction Series, 2016, , 227-261.	0.4	3
66	A probabilistic interpretation for artificial neural network-based voice conversion. , 2015, , .		5
67	Improving denoising auto-encoder based speech enhancement with the speech parameter generation algorithm. , 2015, , .		9
68	Incorporating proximity information in relevance language modeling for extractive speech summarization. , 2015, , .		3
69	A histogram density modeling approach to music emotion recognition. , 2015, , .		10
70	Incorporating paragraph embeddings and density peaks clustering for spoken document summarization. , 2015, , .		4
71	Modeling the Affective Content of Music with a Gaussian Mixture Model. IEEE Transactions on Affective Computing, 2015, 6, 56-68.	5.7	41
72	An Acoustic-Phonetic Model of F0 Likelihood for Vocal Melody Extraction. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 1457-1468.	4.0	4

#	ARTICLE	IF	CITATIONS
73	Extractive Broadcast News Summarization Leveraging Recurrent Neural Network Language Modeling Techniques. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 1322-1334.	4.0	33
74	Combining Relevance Language Modeling and Clarity Measure for Extractive Speech Summarization. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 957-969.	4.0	25
75	I-vector based language modeling for query representation. , 2015, , .		0
76	EMV-matchmaker. , 2015, , .		9
77	A Probabilistic Framework for Chinese Spelling Check. ACM Transactions on Asian and Low-Resource Language Information Processing, 2015, 14, 1-17.	1.3	6
78	A margin-based discriminative modeling approach for extractive speech summarization. , 2014, , .		2
79	Emotion recognition of conversational affective speech using temporal course modeling-based error weighted cross-correlation model. , 2014, , .		3
80	Towards time-varying music auto-tagging based on CAL500 expansion. , 2014, , .		16
81	I-vector based language modeling for spoken document retrieval. , 2014, , .		15
82	Improving music auto-tagging by intra-song instance bagging. , 2014, , .		3
83	Speaker verification using kernel-based binary classifiers with binary operation derived features. , 2014, , .		11
84	A recurrent neural network language modeling framework for extractive speech summarization. , 2014, , .		7
85	Effective pseudo-relevance feedback for language modeling in extractive speech summarization. , 2014, , .		3
86	Generalized <i>k</i> -Labelsets Ensemble for Multi-Label and Cost-Sensitive Classification. IEEE Transactions on Knowledge and Data Engineering, 2014, 26, 1679-1691.	4.0	33
87	Leveraging Effective Query Modeling Techniques for Speech Recognition and Summarization. , 2014, , .		6
88	Non-reference audio quality assessment for online live music recordings. , 2013, , .		16
89	Effective pseudo-relevance feedback for spoken document retrieval. , 2013, , .		2
90	Subspace-based phonotactic language recognition using multivariate dynamic linear models. , 2013, , .		2

#	ARTICLE	IF	CITATIONS
91	Incorporating global variance in the training phase of GMM-based voice conversion. , 2013, , .		13
92	Weighted matrix factorization for spoken document retrieval. , 2013, , .		6
93	Query-Document Relevance Topic Models. Lecture Notes in Computer Science, 2013, , 209-220.	1.0	0
94	Exploring the relationship between categorical and dimensional emotion semantics of music. , 2012, , .		14
95	The acousticvisual emotion guassians model for automatic generation of music video. , 2012, , .		25
96	Generalized k-labelset ensemble for multi-label classification. , 2012, , .		1
97	The acoustic emotion guassians model for emotion-based music annotation and retrieval. , 2012, , .		35
98	Spoken Document Retrieval Leveraging Unsupervised and Supervised Topic Modeling Techniques. IEICE Transactions on Information and Systems, 2012, E95.D, 1195-1205.	0.4	4
99	Playing with tagging: A real-time tagging music player. , 2012, , .		7
100	Exploring mutual information for GMM-based spectral conversion. , 2012, , .		6
101	A Term Association Translation Model for Naive Bayes Text Classification. Lecture Notes in Computer Science, 2012, , 243-253.	1.0	2
102	Automatic annotation of Web videos. , 2011, , .		4
103	Cost-Sensitive Multi-Label Learning for Audio Tag Annotation and Retrieval. IEEE Transactions on Multimedia, 2011, 13, 518-529.	5.2	67
104	Cost-sensitive stacking for audio tag annotation and retrieval. , 2011, , .		3
105	Query by multi-tags with multi-level preferences for content-based music retrieval. , 2011, , .		8
106	Colorizing tags in tag cloud. , 2011, , .		17
107	Audio Tag Annotation and Retrieval Using Tag Count Information. Lecture Notes in Computer Science, 2011, , 339-349.	1.0	1
108	BIC-Based Speaker Segmentation Using Divide-and-Conquer Strategies With Application to Speaker Diarization. IEEE Transactions on Audio Speech and Language Processing, 2010, 18, 141-157.	3.8	27

#	ARTICLE	IF	CITATIONS
109	Background music identification through content filtering and min-hash matching. , 2010, , .		5
110	Detecting pitching frames in baseball game video using Markov random walk. , 2010, , .		2
111	Fast min-hashing indexing and robust spatio-temporal matching for detecting video copies. ACM Transactions on Multimedia Computing, Communications and Applications, 2010, 6, 1-23.	3.0	21
112	Speaker verification using support vector machine with LLR-based sequence kernels. , 2010, , .		1
113	Exploiting semantic associative information in topic modeling. , 2010, , .		4
114	Time-Series Linear Search for Video Copies Based on Compact Signature Manipulation and Containment Relation Modeling. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20, 1603-1613.	5.6	14
115	Homogeneous segmentation and classifier ensemble for audio tag annotation and retrieval. , 2010, , .		11
116	Phone boundary refinement using ranking methods. , 2010, , .		0
117	A Discriminative and Heteroscedastic Linear Feature Transformation for Multiclass Classification. , 2010, , .		1
118	Learning to rank from Bayesian decision inference. , 2009, , .		9
119	A Comparative Study of Probabilistic Ranking Models for Chinese Spoken Document Summarization. ACM Transactions on Asian Language Information Processing, 2009, 8, 1-23.	0.8	23
120	A Probabilistic Generative Framework for Extractive Broadcast News Speech Summarization. IEEE Transactions on Audio Speech and Language Processing, 2009, 17, 95-106.	3.8	23
121	Improving the characterization of the alternative hypothesis via minimum verification error training with applications to speaker verification. Pattern Recognition, 2009, 42, 1351-1360.	5.1	1
122	Evolutionary minimization of the Rand index for speaker clustering. Computer Speech and Language, 2009, 23, 165-175.	2.9	2
123	Improving GMMâ€‘UBM speaker verification using discriminative feedback adaptation. Computer Speech and Language, 2009, 23, 376-388.	2.9	9
124	Model-Based Clustering by Probabilistic Self-Organizing Maps. IEEE Transactions on Neural Networks, 2009, 20, 805-826.	4.8	29
125	Raman-Based 10.66 Gb/s Bidirectional TDM over Long-Reach WDM Hybrid PON. IEICE Transactions on Communications, 2009, E92-B, 3911-3914.	0.4	0
126	Using Kernel Discriminant Analysis to Improve the Characterization of the Alternative Hypothesis for Speaker Verification. IEEE Transactions on Audio Speech and Language Processing, 2008, 16, 1675-1684.	3.8	7



#	ARTICLE	IF	CITATIONS
127	A Query-by-Singing System for Retrieving Karaoke Music. IEEE Transactions on Multimedia, 2008, 10, 1626-1637.	5.2	34
128	Discriminative Feedback Adaptation for GMM-UBM Speaker Verification. , 2008, , .		3
129	A comparative study of probabilistic ranking models for spoken document summarization. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	1
130	An Investigation of Phonological Feature Systems Used in Detection-Based ASR. , 2008, , .		1
131	BIC-based audio segmentation by divide-and-conquer. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	5
132	Automatic Identification of the Sung Language in Popular Music Recordings. Journal of New Music Research, 2007, 36, 105-114.	0.6	6
133	Phonetic Boundary Refinement using Support Vector Machine. , 2007, , .		16
134	Improved Methods for Characterizing the Alternative Hypothesis using Minimum Verification Error Training for LLR-Based Speaker Verification. , 2007, , .		1
135	Cascading Multimodal Verification using Face, Voice and Iris Information. , 2007, , .		8
136	Spoken document summarization using relevant information. , 2007, , .		1
137	Speaker Clustering Based on Minimum Rand Index. , 2007, , .		1
138	Automatic Speaker Clustering Using a Voice Characteristic Reference Space and Maximum Purity Estimation. IEEE Transactions on Audio Speech and Language Processing, 2007, 15, 1461-1474.	3.8	16
139	Integrating coding techniques into LP-based Mandarin text-to-speech synthesis. International Journal of Speech Technology, 2007, 10, 31-44.	1.4	2
140	TONE MODELING FOR SPEECH SYNTHESIS. , 2006, , 77-98.		1
141	SPOKEN DOCUMENT RETRIEVAL AND SUMMARIZATION. , 2006, , 301-320.		2
142	Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals. IEEE Transactions on Audio Speech and Language Processing, 2006, 14, 330-341.	3.8	60
143	A Prototypes-Embedded Genetic K-means Algorithm. , 2006, , .		7
144	Speech utterance clustering based on the maximization of within-cluster homogeneity of speaker voice characteristics. Journal of the Acoustical Society of America, 2006, 120, 1631-1645.	0.5	6

#	ARTICLE	IF	CITATIONS
145	A Kernel-based Discrimination Framework for Solving Hypothesis Testing Problems with Application to Speaker Verification. , 2006, , .		2
146	Extractive Chinese Spoken Document Summarization Using Probabilistic Ranking Models. Lecture Notes in Computer Science, 2006, , 660-671.	1.0	9
147	A Music Retrieval System Based on Query-by-Singing for Karaoke Jukebox. Lecture Notes in Computer Science, 2006, , 445-459.	1.0	0
148	Automatic Construction of Regression Class Tree for MLLR Via Model-Based Hierarchical Clustering. Lecture Notes in Computer Science, 2006, , 390-398.	1.0	0
149	A Minimum Boundary Error Framework for Automatic Phonetic Segmentation. Lecture Notes in Computer Science, 2006, , 399-409.	1.0	4
150	A Novel Alternative Hypothesis Characterization Using Kernel Classifiers for LLR-Based Speaker Verification. Lecture Notes in Computer Science, 2006, , 506-517.	1.0	2
151	Fluent speech prosody: Framework and modeling. Speech Communication, 2005, 46, 284-309.	1.6	97
152	On the extraction of vocal-related information to facilitate the management of popular music collections. , 2005, , .		8
153	A Query-by-Singing Technique for Retrieving Polyphonic Objects of Popular Music. Lecture Notes in Computer Science, 2005, , 439-453.	1.0	5
154	Blind Clustering of Popular Music Recordings Based on Singer Voice Characteristics. Computer Music Journal, 2004, 28, 68-78.	0.3	27
155	A discriminative HMM/N-gram-based retrieval approach for mandarin spoken documents. ACM Transactions on Asian Language Information Processing, 2004, 3, 128-145.	0.8	26
156	The SoVideo Mandarin Chinese Broadcast News Retrieval System. International Journal of Speech Technology, 2004, 7, 189-202.	1.4	5
157	Mandarin-English Information (MEI): investigating translingual speech retrieval. Computer Speech and Language, 2004, 18, 163-179.	2.9	25
158	A Model-Selection-Based Self-Splitting Gaussian Mixture Learning with Application to Speaker Identification. Eurasip Journal on Advances in Signal Processing, 2004, 2004, 1.	1.0	13
159	Discriminating capabilities of syllable-based features and approaches of utilizing them for voice retrieval of speech information in Mandarin Chinese. IEEE Transactions on Speech and Audio Processing, 2002, 10, 303-314.	2.0	45
160	A hierarchical tag-graph search scheme with layered grammar rules for spontaneous speech understanding. Pattern Recognition Letters, 2002, 23, 819-831.	2.6	6
161	Content-based Language Models for Spoken Document Retrieval. International Journal of Computer Processing of Languages, 2001, 14, 193-209.	0.3	1
162	Mandarin-English Information (MEI). , 2001, , .		10

#	ARTICLE	IF	CITATIONS
163	Comparison of Word and Subword Indexing Techniques for Mandarin Chinese Spoken Document Retrieval. Lecture Notes in Computer Science, 2001, , 606-613.	1.0	1
164	A spoken-access approach for chinese text and speech information retrieval. Journal of the Association for Information Science and Technology, 2000, 51, 313-323.	1.2	5
165	Experiments in syllable-based retrieval of broadcast news speech in Mandarin Chinese. Speech Communication, 2000, 32, 49-60.	1.6	27
166	Mandarin spoken document retrieval based on syllable lattice matching. Pattern Recognition Letters, 2000, 21, 615-624.	2.6	1
167	Browsing the Chinese Web Pages Using Mandarin Speech. International Journal of Computer Processing of Languages, 2000, 13, 35-51.	0.3	1
168	Content-based language models for spoken document retrieval. , 2000, , .		2
169	SYLLABLE-BASED CHINESE TEXT/SPOKEN DOCUMENT RETRIEVAL USING TEXT/SPEECH QUERIES. International Journal of Pattern Recognition and Artificial Intelligence, 2000, 14, 603-616.	0.7	9
170	Mandarin-English Information (MEI). , 2000, , .		10
171	Automatic selection of phonetically distributed sentence sets for speaker adaptation with application to large vocabulary Mandarin speech recognition. Computer Speech and Language, 1999, 13, 79-97.	2.9	15
172	Surfing The Chinese Web Pages By Unconstrained Mandarin Speech. , 1998, , .		1
173	A Prototype Of Mandarin Voice Memo System. , 1998, , .		0
174	Complete recognition of continuous Mandarin speech for Chinese language with very large vocabulary using limited training data. IEEE Transactions on Speech and Audio Processing, 1997, 5, 195-200.	2.0	38
175	Frameworks for recognition of Mandarin syllables with tones using sub-syllabic units. Speech Communication, 1996, 18, 175-190.	1.6	20
176	An initial study on a segmental probability model approach to large-vocabulary continuous Mandarin speech recognition. , 0, , .		3
177	Complete recognition of continuous Mandarin speech for Chinese language with very large vocabulary but limited training data. , 0, , .		18
178	Internet Chinese information retrieval using unconstrained Mandarin speech queries based on a client-server architecture and a PAT-tree-based language model. , 0, , .		2
179	A key-phrase understanding framework integrating real world knowledge with speech recognition with initial application in voice memo systems for Mandarin Chinese. , 0, , .		2
180	A word-length-dependent confidence measure for large vocabulary Chinese keyword spotting. , 0, , .		0

#	ARTICLE	IF	CITATIONS
181	Retrieval of broadcast news speech in Mandarin Chinese collected in Taiwan using syllable-level statistical characteristics. , 0, , .		23
182	Multi-scale-audio indexing for translingual spoken document retrieval. , 0, , .		9
183	Eigenspace-based maximum a posteriori linear regression for rapid speaker adaptation. , 0, , .		12
184	A new eigenvoice approach to speaker adaptation. , 0, , .		5
185	A Mandarin TTS system with an integrated prosodic model. , 0, , .		2
186	Automatic detection and tracking of target singer in multi-singer music recordings. , 0, , .		5
187	A query-by-example framework to retrieve music documents by singer. , 0, , .		1
188	A maximum entropy approach for integrating semantic information in statistical language models. , 0, , .		4
189	Gmm-Based Bhattacharyya Kernel Fisher Discriminant Analysis For Speaker Recognition. , 0, , .		3
190	Clustering Speech Utterances by Speaker Using Eigenvoice-Motivated Vector Space Models. , 0, , .		9
191	An Efficient Approach to Multimodal Person Identity Verification by Fusing Face and Voice Information. , 0, , .		6
192	On Maximizing the Within-Cluster Homogeneity of Speaker Voice Characteristics For Speech Utterance Clustering. , 0, , .		2
193	Locally Linear Embedding for Exemplar-Based Spectral Conversion. , 0, , .		20
194	Exploring Word Moverâ€™s Distance and Semantic-Aware Embedding Techniques for Extractive Broadcast News Summarization. , 0, , .		4
195	Minimization of Regression and Ranking Losses with Shallow Neural Networks on Automatic Sincerity Evaluation. , 0, , .		2
196	Exploring the Use of Significant Words Language Modeling for Spoken Document Retrieval. , 0, , .		5
197	A Post-Filtering Approach Based on Locally Linear Embedding Difference Compensation for Speech Enhancement. , 0, , .		2
198	Voice Conversion from Unaligned Corpora Using Variational Autoencoding Wasserstein Generative Adversarial Networks. , 0, , .		125

#	ARTICLE	IF	CITATIONS
199	Wavelet Speech Enhancement Based on Robust Principal Component Analysis. , 0, , .		2
200	Quality-Net: An End-to-End Non-intrusive Speech Quality Assessment Model Based on BLSTM. , 0, , .		75
201	Investigation of F0 Conditioning and Fully Convolutional Networks in Variational Autoencoder Based Voice Conversion. , 0, , .		5
202	MOSNet: Deep Learning-Based Objective Assessment for Voice Conversion. , 0, , .		92
203	Speaker clustering of unknown utterances based on maximum purity estimation. , 0, , .		5
204	Alleviating the over-smoothing problem in GMM-based voice conversion with discriminative training. , 0, , .		4
205	Ensemble of machine learning algorithms for cognitive and physical speaker load detection. , 0, , .		2
206	Clustering-based i-vector formulation for speaker recognition. , 0, , .		1
207	Discriminative Autoencoders for Acoustic Modeling. , 0, , .		3
208	Exemplar-Based Spectral Detail Compensation for Voice Conversion. , 0, , .		2
209	Exploring the Encoder Layers of Discriminative Autoencoders for LVCSR. , 0, , .		4
210	Specialized Speech Enhancement Model Selection Based on Learned Non-Intrusive Quality Assessment Metric. , 0, , .		6
211	Noise Adaptive Speech Enhancement Using Domain Adversarial Training. , 0, , .		18
212	Lite Audio-Visual Speech Enhancement. , 0, , .		11