

Helen M Meng

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/4515236/publications.pdf>

Version: 2024-02-01

165
papers

2,256
citations

567144

15
h-index

501076

28
g-index

170
all docs

170
docs citations

170
times ranked

1081
citing authors

#	ARTICLE	IF	CITATIONS
1	Creation and Evaluation of a Pretertiary Artificial Intelligence (AI) Curriculum. IEEE Transactions on Education, 2022, 65, 30-39.	2.0	72
2	Any-to-Many Voice Conversion With Location-Relative Sequence-to-Sequence Modeling. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 1717-1728.	4.0	24
3	Audio-Visual Multi-Channel Integration and Recognition of Overlapped Speech. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 2067-2082.	4.0	14
4	Exemplar-Based Emotive Speech Synthesis. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 874-886.	4.0	8
5	Automatic Speaker-level Pronunciation Assessment of L2 Speech Using Posterior Probabilities from Multiple Utterances. , 2021, , .		1
6	Adversarial Defense for Automatic Speaker Verification by Cascaded Self-Supervised Learning Models. , 2021, , .		18
7	Fcl-Taco2: Towards Fast, Controllable and Lightweight Text-to-Speech Synthesis. , 2021, , .		6
8	Bayesian Learning of LF-MMI Trained Time Delay Neural Networks for Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 1514-1529.	4.0	7
9	Speech Emotion Recognition Using Sequential Capsule Networks. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 3280-3291.	4.0	14
10	DiffSVC: A Diffusion Probabilistic Model for Singing Voice Conversion. , 2021, , .		8
11	Reconstructing Dual Learning for Neural Voice Conversion Using Relatively Few Samples. , 2021, , .		6
12	End-To-End Voice Conversion Via Cross-Modal Knowledge Distillation for Dysarthric Speech Reconstruction. , 2020, , .		13
13	Adversarial Attacks on GMM I-Vector Based Speaker Verification Systems. , 2020, , .		41
14	End-To-End Accent Conversion Without Using Native Utterances. , 2020, , .		14
15	An Integrated Approach of Machine Learning and Systems Thinking for Waiting Time Prediction in an Emergency Department. International Journal of Medical Informatics, 2020, 139, 104143.	1.6	34
16	Code-Switched Speech Synthesis Using Bilingual Phonetic Posteriorgram with Only Monolingual Corpora. , 2020, , .		11
17	CNN-RNN-CTC Based End-to-end Mispronunciation Detection and Diagnosis. , 2019, , .		38
18	Adversarial Attacks on Spoofing Countermeasures of Automatic Speaker Verification. , 2019, , .		26

#	ARTICLE	IF	CITATIONS
19	Learning Contextual Representation with Convolution Bank and Multi-head Self-attention for Speech Emphasis Detection. , 2019, , .		0
20	Towards Discriminative Representation Learning for Speech Emotion Recognition. , 2019, , .		26
21	Automatic lexical stress and pitch accent detection for L2 English speech using multi-distribution deep neural networks. Speech Communication, 2018, 96, 28-36.	1.6	37
22	Integrating Articulatory Features into Acoustic-Phonemic Model for Mispronunciation Detection and Diagnosis in L2 English Speech. , 2018, , .		7
23	Topic Discovery via Convex Polytopic Model: A Case Study with Small Corpora. , 2018, , .		1
24	Unsupervised Discovery of an Extended Phoneme Set in L2 English Speech for Mispronunciation Detection and Diagnosis. , 2018, , .		7
25	DNN i-vector based Fishvoice and PLDA SVM scoring for NIST SRE 2016. , 2018, , .		0
26	Applying Multitask Learning to Acoustic-Phonemic Model for Mispronunciation Detection and Diagnosis in L2 English Speech. , 2018, , .		8
27	Emphatic Speech Generation with Conditioned Input Layer and Bidirectional LSTMS for Expressive Speech Synthesis. , 2018, , .		7
28	Limited-Memory BFGS Optimization of Recurrent Neural Network Language Models for Speech Recognition. , 2018, , .		10
29	Using simulation to assess the impacts of the adoption of a fast-track system for hospital emergency services. Journal of Advanced Mechanical Design, Systems and Manufacturing, 2018, 12, JAMDSM0073-JAMDSM0073.	0.3	7
30	Intonation classification for L2 English speech using multi-distribution deep neural networks. Computer Speech and Language, 2017, 43, 18-33.	2.9	21
31	Parallel probabilistic swarm guidance by exploiting Kronecker product structures in discrete-time Markov chains. , 2017, , .		0
32	Learning cross-lingual knowledge with multilingual BLSTM for emphasis detection with limited training data. , 2017, , .		12
33	A model of extended paragraph vector for document categorization and trend analysis. , 2017, , .		1
34	Multi-task learning of structured output layer bidirectional LSTMS for speech synthesis. , 2017, , .		7
35	Mispronunciation Detection and Diagnosis in L2 English Speech Using Multidistribution Deep Neural Networks. IEEE/ACM Transactions on Audio Speech and Language Processing, 2017, 25, 193-207.	4.0	88
36	Exploring articulatory characteristics of Cantonese dysarthric speech using distinctive features. , 2016, , .		2

#	ARTICLE	IF	CITATIONS
37	DBLSTM-based multi-task learning for pitch transformation in voice conversion. , 2016, , .		1
38	Kronecker product approximation with multiple factor matrices via the tensor product algorithm. , 2016, , .		3
39	Phonetic posteriorgrams for many-to-one voice conversion without parallel data training. , 2016, , .		149
40	An embedding approach for context-aware collaborative recommendation and visualization. , 2016, , .		1
41	Recognizing stances in Mandarin social ideological debates with text and acoustic features. , 2016, , .		0
42	Low level descriptors based DBLSTM bottleneck feature for speech driven talking avatar. , 2016, , .		4
43	Question detection from acoustic features using recurrent neural network with gated recurrent unit. , 2016, , .		33
44	Exploratory data analysis on nuclei in cantonese dysarthric speech. , 2016, , .		0
45	Learning cross-lingual information with multilingual BLSTM for speech synthesis of low-resource languages. , 2016, , .		19
46	Learning Track Representation and Trends for Conference Analytics. , 2016, , .		1
47	Capitalizing on musical rhythm for prosodic training in computer-aided language learning. Computer Speech and Language, 2016, 37, 67-81.	2.9	12
48	A Two-Pass Framework of Mispronunciation Detection and Diagnosis for Computer-Aided Pronunciation Training. IEEE/ACM Transactions on Audio Speech and Language Processing, 2016, 24, 1020-1028.	4.0	20
49	Understanding speaking styles of internet speech data with LSTM and low-resource training. , 2015, , .		0
50	A two-pass framework of mispronunciation detection & diagnosis for computer-aided pronunciation training. , 2015, , .		2
51	Acoustic to articulatory mapping with deep neural network. Multimedia Tools and Applications, 2015, 74, 9889-9907.	2.6	17
52	HMM-based emphatic speech synthesis for corrective feedback in computer-aided pronunciation training. , 2015, , .		7
53	AA spectral space warping approach to cross-lingual voice transformation in HMM-based TTS. , 2015, , .		8
54	Voice conversion using deep Bidirectional Long Short-Term Memory based Recurrent Neural Networks. , 2015, , .		162

#	ARTICLE	IF	CITATIONS
55	A Data Capturing Platform in the Cloud for Behavioral Analysis among Smokers: An Application Platform for Public Health Research. , 2015, , .		2
56	A deep recurrent approach for acoustic-to-articulatory inversion. , 2015, , .		43
57	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.	4.0	1
58	Generating emphatic speech with hidden Markov model for expressive speech synthesis. Multimedia Tools and Applications, 2015, 74, 9909-9925.	2.6	8
59	Expressive talking avatar synthesis and animation. Multimedia Tools and Applications, 2015, 74, 9845-9848.	2.6	5
60	Deep Learning for Acoustic Modeling in Parametric Speech Generation: A systematic review of existing techniques and future trends. IEEE Signal Processing Magazine, 2015, 32, 35-52.	4.6	177
61	Blood Pressure Management with Data Capturing in the Cloud among Hypertensive Patients: A Monitoring Platform for Hypertensive Patients. , 2015, , .		2
62	Indoor Air Monitoring Platform and Personal Health Reporting System: Big Data Analytics for Public Health Research. , 2015, , .		10
63	Embracing Big Data for Simulation Modelling of Emergency Department Processes and Activities. , 2015, , .		8
64	A Real-Time Decision Support Tool for Disaster Response: A Mathematical Programming Approach. , 2015, , .		8
65	Analysis of Dysarthric Speech using Distinctive Feature Recognition. , 2015, , .		5
66	Grading the Severity of Mispronunciations in CAPT Based on Statistical Analysis and Computational Speech Perception. Journal of Computer Science and Technology, 2014, 29, 751-761.	0.9	2
67	Phonological modeling of mispronunciation gradations in L2 English speech of L1 Chinese learners. , 2014, , .		7
68	Automatic detection of contrastive word pairs using textual and acoustic features. , 2014, , .		0
69	Automatic speech data clustering with human perception based weighted distance. , 2014, , .		3
70	Mispronunciation detection and diagnosis in l2 english speech using multi-distribution Deep Neural Networks. , 2014, , .		5
71	Learning dynamic features with neural networks for phoneme recognition. , 2014, , .		2
72	Contrastive auto-encoder for phoneme recognition. , 2014, , .		5

#	ARTICLE	IF	CITATIONS
73	Synthesizing English emphatic speech for multimodal corrective feedback in computer-aided pronunciation training. <i>Multimedia Tools and Applications</i> , 2014, 73, 463-489.	2.6	18
74	Head and facial gestures synthesis using PAD model for an expressive talking avatar. <i>Multimedia Tools and Applications</i> , 2014, 73, 439-461.	2.6	23
75	Latent Semantic Analysis for Multimodal User Input With Speech and Gestures. <i>IEEE/ACM Transactions on Audio Speech and Language Processing</i> , 2014, 22, 417-429.	4.0	13
76	Multi-distribution deep belief network for speech synthesis. , 2013, , .		74
77	Predicting gradation of L2 English mispronunciations using ASR with extended recognition network. , 2013, , .		1
78	Development of text-to-audiovisual speech synthesis to support interactive language learning on a mobile device. , 2013, , .		5
79	Audiovisual synthesis of exaggerated speech for corrective feedback in computer-assisted pronunciation training. , 2013, , .		8
80	Investigation of tandem deep belief network approach for phoneme recognition. , 2013, , .		2
81	Clustering similar acoustic classes in the Fishervoice framework. , 2013, , .		4
82	Generating emphasis from neutral speech using hierarchical perturbation model by decision tree and support vector machine. , 2012, , .		0
83	Deriving perceptual gradation OF L2 English mispronunciations using crowdsourcing and the WorkerRank algorithm. , 2012, , .		4
84	mENUNCIATE: Development of a computer-aided pronunciation training system on a cross-platform framework for mobile, speech-enabled application development. , 2012, , .		7
85	Detection and emphatic realization of contrastive word pairs for expressive text-to-speech synthesis. , 2012, , .		2
86	Predicting User Satisfaction in Spoken Dialog System Evaluation With Collaborative Filtering. <i>IEEE Journal on Selected Topics in Signal Processing</i> , 2012, 6, 971-981.	7.3	23
87	Perceptually-motivated assessment of automatically detected lexical stress in L2 learners' speech. , 2012, , .		3
88	Analysis on mispronunciations in CAPT based on computational speech perception. , 2012, , .		0
89	Phoneme-level articulatory animation in pronunciation training. <i>Speech Communication</i> , 2012, 54, 845-856.	1.6	39
90	Farewell Editorial. <i>IEEE Transactions on Audio Speech and Language Processing</i> , 2012, 20, 1-1.	3.8	4

#	ARTICLE	IF	CITATIONS
91	Enunciate: An internet-accessible computer-aided pronunciation training system and related user evaluations. , 2011, , .		9
92	Allophonic variations in visual speech synthesis for corrective feedback in CAPT. , 2011, , .		12
93	The HKCUPU system for the NIST 2010 speaker recognition evaluation. , 2011, , .		1
94	Pseudo-Conventional N-Gram Representation of the Discriminative N-Gram Model for LVCSR. IEEE Journal on Selected Topics in Signal Processing, 2010, 4, 943-952.	7.3	2
95	Introduction to the Issue on Statistical Learning Methods for Speech and Language Processing. IEEE Journal on Selected Topics in Signal Processing, 2010, 4, 913-916.	7.3	0
96	Statistical phone duration modeling to filter for intact utterances in a computer-assisted pronunciation training system. , 2010, , .		7
97	Collection of user judgments on spoken dialog system with crowdsourcing. , 2010, , .		23
98	Fishervoice: A discriminant subspace framework for speaker recognition. , 2010, , .		5
99	Usage patterns and latent semantic analyses for task goal inference of multimodal user interactions. , 2010, , .		1
100	Modeling prosody patterns for Chinese expressive text-to-speech synthesis. , 2010, , .		1
101	Using finite state machines for evaluating spoken dialog systems. , 2010, , .		3
102	Collaborative filtering model for user satisfaction prediction in Spoken Dialog System evaluation. , 2010, , .		8
103	Detection of intonation in L2 English speech of native Mandarin learners. , 2010, , .		5
104	Capturing L2 segmental mispronunciations with joint-sequence models in Computer-Aided Pronunciation Training (CAPT). , 2010, , .		18
105	Predicting user evaluations of spoken dialog systems using semi-supervised learning. , 2010, , .		2
106	An enhanced Fishervoice subspace framework for text-independent speaker verification. , 2010, , .		6
107	Development of an articulatory visual-speech synthesizer to support language learning. , 2010, , .		7
108	Audiovisual Tools for Phonetic and Articulatory Visualization in Computer-Aided Pronunciation Training. Lecture Notes in Computer Science, 2010, , 337-345.	1.0	8

#	ARTICLE	IF	CITATIONS
109	Facial Expression Synthesis Based on Emotion Dimensions for Affective Talking Avatar. Smart Innovation, Systems and Technologies, 2010, , 109-132.	0.5	13
110	Cross-Modality Semantic Integration With Hypothesis Rescoring for Robust Interpretation of Multimodal User Interactions. IEEE Transactions on Audio Speech and Language Processing, 2009, 17, 486-500.	3.8	2
111	Introduction to the Special Issue on Multimodal Processing in Speech-Based Interactions. IEEE Transactions on Audio Speech and Language Processing, 2009, 17, 409-410.	3.8	2
112	Modeling the Expressivity of Input Text Semantics for Chinese Text-to-Speech Synthesis in a Spoken Dialog System. IEEE Transactions on Audio Speech and Language Processing, 2009, 17, 1567-1576.	3.8	13
113	Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). , 2009, , .		23
114	The Use of Dynamic Deformable Templates for Lip Tracking in an Audio-Visual Corpus with Large Variations in Head Pose, Face Illumination and Lip Shapes. , 2008, , .		2
115	A New Prosodic Strength Calculation Method for Prosody Reduction Modeling. , 2008, , .		2
116	Recasting the discriminative n-gram model as a pseudo-conventional n-gram model for LVCSR. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, , .	1.8	4
117	Mispronunciation detection based on cross-language phonological comparisons. , 2008, , .		7
118	Decision Fusion for Improving Mispronunciation Detection Using Language Transfer Knowledge and Phoneme-Dependent Pronunciation Scoring. , 2008, , .		5
119	Adaptive Weight Estimation in Multi-Biometric Verification using Fuzzy Logic Decision Fusion. , 2007, , .		9
120	Discriminant Mutual Subspace Learning for Indoor and Outdoor Face Recognition. , 2007, , .		2
121	Speaker Verification via High-Level Feature Based Phonetic-Class Pronunciation Modeling. IEEE Transactions on Computers, 2007, 56, 1189-1198.	2.4	11
122	Deriving salient learners' mispronunciations from cross-language phonological comparisons. , 2007, , .		28
123	Head Movement Synthesis Based on Semantic and Prosodic Features for a Chinese Expressive Avatar. , 2007, , .		12
124	Effects of Device Mismatch, Language Mismatch and Environmental Mismatch on Speaker Verification. , 2007, , .		4
125	Facial Expression Synthesis Using PAD Emotional Parameters for a Chinese Expressive Avatar. Lecture Notes in Computer Science, 2007, , 24-35.	1.0	18
126	MODELLING THE GLOBAL ACOUSTIC CORRELATES OF EXPRESSIVITY FOR CHINESE TEXT-TO-SPEECH SYNTHESIS. , 2006, , .		9

#	ARTICLE	IF	CITATIONS
127	MULTILINGUAL DIALOG SYSTEMS. , 2006, , 459-481.		0
128	Initial Experiments on Automatic Story Segmentation in Chinese Spoken Documents Using Lexical Cohesion of Extracted Named Entities. Lecture Notes in Computer Science, 2006, , 693-703.	1.0	3
129	CANTONESE SPEECH RECOGNITION AND SYNTHESIS. , 2006, , 365-386.		10
130	Multilingual Spoken Dialog Systems. , 2006, , 399-447.		3
131	A Corpus-Based Approach for Cooperative Response Generation in a Dialog System. Lecture Notes in Computer Science, 2006, , 614-626.	1.0	4
132	A Cantonese Speech-Driven Talking Face Using Translingual Audio-to-Visual Conversion. Lecture Notes in Computer Science, 2006, , 627-639.	1.0	0
133	Using Verb Dependency Matching in a Reading Comprehension System. Lecture Notes in Computer Science, 2005, , 190-201.	1.0	0
134	A Pruning Approach for GMM-Based Speaker Verification in Mobile Embedded Systems. Lecture Notes in Computer Science, 2004, , 607-613.	1.0	3
135	Mandarin-English Information (MEI): investigating translingual speech retrieval. Computer Speech and Language, 2004, 18, 163-179.	2.9	25
136	ISIS. ACM Transactions on Computer-Human Interaction, 2004, 11, 268-299.	4.6	9
137	The State of the Art in Human-computer Speech-based Interface Technologies. HKIE Transactions, 2003, 10, 50-61.	1.9	0
138	Intelligent speech for information systems: towards biliteracy and trilingualism. Interacting With Computers, 2002, 14, 327-339.	1.0	2
139	Spoken language resources for Cantonese speech processing. Speech Communication, 2002, 36, 327-342.	1.6	76
140	Interdependencies among dialog acts, task goals and discourse inheritance in mixed-initiative dialogs. , 2002, , .		2
141	A hierarchical lexical representation for bi-directional spelling-to-pronunciation/pronunciation-to-spelling generation. Speech Communication, 2001, 33, 213-239.	1.6	1
142	A Hierarchical Lexical Representation for Pronunciation Generation. , 2001, , 71-90.		0
143	Initial Development Towards a Trilingual Speech Interface for Financial Information Inquiries. International Journal of Speech Technology, 2000, 3, 83-91.	1.4	2
144	Mandarin-English Information (MEI). , 2000, , .		10

#	ARTICLE	IF	CITATIONS
145	Reversible letter-to-sound/sound-to-letter generation based on parsing word morphology. <i>Speech Communication</i> , 1996, 18, 47-63.	1.6	12
146	Real-time synthesis of Chinese visual speech and facial expressions using MPEG-4 FAP features in a three-dimensional avatar. , 0, , .		25
147	Modeling the acoustic correlates of expressive elements in text genres for expressive text-to-speech synthesis. , 0, , .		5
148	High-level feature-based speaker verification via articulatory phonetic-class pronunciation modeling. , 0, , .		0
149	Complementarity and redundancy in multimodal user inputs with speech and pen gestures. , 0, , .		1
150	Modeling the statistical behavior of lexical chains to capture word cohesiveness for automatic story segmentation. , 0, , .		14
151	Improving mispronunciation detection and diagnosis of learners' speech with context-sensitive phonological rules based on language transfer. , 0, , .		21
152	Automatic generation and pruning of phonetic mispronunciations to support computer-aided pronunciation training. , 0, , .		9
153	Studying L2 suprasegmental features in asian Englishes: a position paper. , 0, , .		21
154	Automatic derivation of phonological rules for mispronunciation detection in a computer-assisted pronunciation training system. , 0, , .		35
155	Discriminative acoustic model for improving mispronunciation detection and diagnosis in computer-aided pronunciation training (CAPT). , 0, , .		23
156	Prominence model for prosodic features in automatic lexical stress and pitch accent detection. , 0, , .		8
157	An analysis framework based on random subspace sampling for speaker verification. , 0, , .		6
158	On mispronunciation lexicon generation using joint-sequence multigrams in computer-aided pronunciation training (CAPT). , 0, , .		6
159	The use of DBN-HMMs for mispronunciation detection and diagnosis in L2 English to support computer-aided pronunciation training. , 0, , .		27
160	Hierarchical English emphatic speech synthesis based on HMM with limited training data. , 0, , .		3
161	Lexical stress detection for L2 English speech using deep belief networks. , 0, , .		10
162	Using conditional random fields to predict focus word pair in spontaneous spoken English. , 0, , .		1

#	ARTICLE	IF	CITATIONS
163	Development of a Cantonese dysarthric speech corpus. , 0, , .		4
164	Topic modeling for conference analytics. , 0, , .		3
165	Using tilt for automatic emphasis detection with Bayesian networks. , 0, , .		3