

Kai Yu

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/5795080/publications.pdf>

Version: 2024-02-01

112
papers

2,461
citations

471509

17
h-index

330143

37
g-index

112
all docs

112
docs citations

112
times ranked

1472
citing authors

#	ARTICLE	IF	CITATIONS
1	Data augmentation based non-parallel voice conversion with frame-level speaker disentangler. <i>Speech Communication</i> , 2022, 136, 14-22.	2.8	6
2	Phone-Level Prosody Modelling With GMM-Based MDN for Diverse and Controllable Speech Synthesis. <i>IEEE/ACM Transactions on Audio Speech and Language Processing</i> , 2022, 30, 190-201.	5.8	6
3	Neural Fusion for Voice Cloning. <i>IEEE/ACM Transactions on Audio Speech and Language Processing</i> , 2022, 30, 1993-2001.	5.8	3
4	Speech Enhancement with Neural Homomorphic Synthesis. , 2022, , .		6
5	Diversity-Controllable and Accurate Audio Captioning Based on Neural Condition. , 2022, , .		2
6	Navigating Audio-Visual Event Detection Across Mismatched Modalities. , 2022, , .		0
7	Unsupervised Word-Level Prosody Tagging for Controllable Speech Synthesis. , 2022, , .		5
8	LatticeBART: Lattice-to-Lattice Pre-Training for Speech Recognition. , 2022, , .		0
9	Category-Adapted Sound Event Enhancement with Weakly Labeled Data. , 2022, , .		2
10	Text Adaptive Detection for Customizable Keyword Spotting. , 2022, , .		4
11	Climate and Weather: Inspecting Depression Detection via Emotion Recognition. , 2022, , .		5
12	Audio-Text Retrieval in Context. , 2022, , .		9
13	Revisiting the Statistics Pooling Layer in Deep Speaker Embedding Learning. , 2021, , .		11
14	Audio Caption in a Car Setting with a Sentence-Level Loss. , 2021, , .		4
15	Text-to-Audio Grounding: Building Correspondence Between Captions and Sound Events. , 2021, , .		8
16	SynAug: Synthesis-Based Data Augmentation for Text-Dependent Speaker Verification. , 2021, , .		8
17	Investigating Local and Global Information for Automated Audio Captioning with Transfer Learning. , 2021, , .		16
18	Voice Activity Detection in the Wild: A Data-Driven Approach Using Teacher-Student Training. <i>IEEE/ACM Transactions on Audio Speech and Language Processing</i> , 2021, 29, 1542-1555.	5.8	19

#	ARTICLE	IF	CITATIONS
19	Towards Duration Robust Weakly Supervised Sound Event Detection. IEEE/ACM Transactions on Audio Speech and Language Processing, 2021, 29, 887-900.	5.8	25
20	The practice of speech and language processing in China. Communications of the ACM, 2021, 64, 81-87.	4.5	0
21	Modular End-to-End Automatic Speech Recognition Framework for Acoustic-to-Word Model. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 2174-2183.	5.8	6
22	Neural Network Language Model Compression With Product Quantization and Soft Binarization. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 2438-2449.	5.8	4
23	Distributed Structured Actor-Critic Reinforcement Learning for Universal Dialogue Management. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 2400-2411.	5.8	10
24	Data Augmentation Using Deep Generative Models for Embedding Based Speaker Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, 28, 2598-2609.	5.8	30
25	<p>Individual Perceived Stress Mediates Psychological Distress in Medical Workers During COVID-19 Epidemic Outbreak in Wuhan</p>. Neuropsychiatric Disease and Treatment, 2020, Volume 16, 2529-2537.	2.2	18
26	A Hierarchical Tracker for Multi-Domain Dialogue State Tracking. , 2020, , .		2
27	Investigation of Specaugment for Deep Speaker Embedding Learning. , 2020, , .		18
28	Speaker Augmentation for Low Resource Speech Recognition. , 2020, , .		20
29	Towards a new generation of artificial intelligence in China. Nature Machine Intelligence, 2020, 2, 312-316.	16.0	90
30	Dual Learning for Semi-Supervised Natural Language Understanding. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, , 1-1.	5.8	15
31	Duration Robust Weakly Supervised Sound Event Detection. , 2020, , .		15
32	Channel Invariant Speaker Embedding Learning with Joint Multi-Task and Adversarial Training. , 2020, , .		8
33	Prior Knowledge Driven Label Embedding for Slot Filling in Natural Language Understanding. IEEE/ACM Transactions on Audio Speech and Language Processing, 2020, , 1-1.	5.8	2
34	Text Adaptation for Speaker Verification with Speaker-Text Factorized Embeddings. , 2020, , .		4
35	Memory Attention Neural Network for Multi-domain Dialogue State Tracking. Lecture Notes in Computer Science, 2020, , 41-52.	1.3	0
36	Discriminative Neural Embedding Learning for Short-Duration Text-Independent Speaker Verification. IEEE/ACM Transactions on Audio Speech and Language Processing, 2019, 27, 1686-1696.	5.8	26

#	ARTICLE	IF	CITATIONS
37	End-to-end Monaural Multi-speaker ASR System without Pretraining. , 2019, , .		39
38	AgentGraph: Toward Universal Dialogue Management With Structured Deep Reinforcement Learning. IEEE/ACM Transactions on Audio Speech and Language Processing, 2019, 27, 1378-1391.	5.8	25
39	A Hierarchical Decoding Model for Spoken Language Understanding from Unaligned Data. , 2019, , .		9
40	Knowledge Distillation for Small Foot-print Deep Speaker Embedding. , 2019, , .		15
41	Audio Caption: Listen and Tell. , 2019, , .		20
42	Margin Matters: Towards More Discriminative Deep Neural Network Embeddings for Speaker Recognition. , 2019, , .		68
43	Highly Efficient Neural Network Language Model Compression Using Soft Binarization Training. , 2019, , .		5
44	Rich Short Text Conversation Using Semantic-Key-Controlled Sequence Generation. IEEE/ACM Transactions on Audio Speech and Language Processing, 2018, 26, 1359-1368.	5.8	7
45	Adaptive Very Deep Convolutional Residual Network for Noise Robust Speech Recognition. IEEE/ACM Transactions on Audio Speech and Language Processing, 2018, 26, 1393-1405.	5.8	52
46	Robust Spoken Language Understanding with Unsupervised ASR-Error Adaptation. , 2018, , .		12
47	Policy Adaptation for Deep Reinforcement Learning-Based Dialogue Management. , 2018, , .		15
48	Fast Oov Words Incorporation Using Structured Word Embeddings for Neural Network Language Model. , 2018, , .		3
49	Semi-Supervised Training Using Adversarial Multi-Task Learning for Spoken Language Understanding. , 2018, , .		11
50	Generative Adversarial Networks based X-vector Augmentation for Robust Probabilistic Linear Discriminant Analysis in Speaker Verification. , 2018, , .		11
51	Deep Discriminant Analysis for i-vector Based Robust Speaker Recognition. , 2018, , .		5
52	Covariance Based Deep Feature for Text-Dependent Speaker Verification. Lecture Notes in Computer Science, 2018, , 231-242.	1.3	2
53	On Modular Training of Neural Acoustics-to-Word Model for LVCSR. , 2018, , .		23
54	Focal KL-Divergence Based Dilated Convolutional Neural Networks for Co-Channel Speaker Identification. , 2018, , .		7

#	ARTICLE	IF	CITATIONS
55	Sequence discriminative training for deep learning based acoustic keyword spotting. <i>Speech Communication</i> , 2018, 102, 100-111.	2.8	11
56	Investigating Raw Wave Deep Neural Networks for End-to-End Speaker Spoofing Detection. <i>IEEE/ACM Transactions on Audio Speech and Language Processing</i> , 2018, 26, 2002-2014.	5.8	38
57	Experimental Study on the Effects of Nozzle Temperature on Internal Deposits of a Gasoline Direct Injector. <i>Energy & Fuels</i> , 2018, 32, 8978-8985.	5.1	5
58	Phone Synchronous Speech Recognition With CTC Lattices. <i>IEEE/ACM Transactions on Audio Speech and Language Processing</i> , 2017, 25, 90-101.	5.8	19
59	Encoder-decoder with focus-mechanism for sequence labelling based spoken language understanding. , 2017, , .		33
60	End-to-end spoofing detection with raw waveform CLDNNS. , 2017, , .		31
61	Small-footprint convolutional neural network for spoofing detection. , 2017, , .		2
62	Confidence measures for CTC-based phone synchronous decoding. , 2017, , .		5
63	Affordable On-line Dialogue Policy Learning. , 2017, , .		9
64	Agent-Aware Dropout DQN for Safe and Efficient On-line Dialogue Policy Learning. , 2017, , .		19
65	On-line Dialogue Policy Learning with Companion Teaching. , 2017, , .		8
66	Multi-task joint-learning for robust voice activity detection. , 2016, , .		2
67	An investigation on deep learning with beta stabilizer. , 2016, , .		0
68	Rich punctuations prediction using large-scale deep learning. , 2016, , .		2
69	Evolvable dialogue state tracking for statistical dialogue management. <i>Frontiers of Computer Science</i> , 2016, 10, 201-215.	2.4	1
70	Very Deep Convolutional Neural Networks for Noise Robust Speech Recognition. <i>IEEE/ACM Transactions on Audio Speech and Language Processing</i> , 2016, 24, 2263-2276.	5.8	262
71	Deep features for automatic spoofing detection. <i>Speech Communication</i> , 2016, 85, 43-52.	2.8	50
72	Phone-aware LSTM-RNN for voice conversion. , 2016, , .		7

#	ARTICLE	IF	CITATIONS
73	Directed automatic speech transcription error correction using bidirectional LSTM. , 2016, , .		7
74	A comparative study of robustness of deep learning approaches for VAD. , 2016, , .		29
75	Discriminatively trained joint speaker and environment representations for adaptation of deep neural network acoustic models. , 2016, , .		2
76	A novel static parameter calculation method for model compensation. , 2015, , .		0
77	Recurrent neural network language model with structured word embeddings for speech recognition. , 2015, , .		2
78	Multi-task joint-learning of deep neural networks for robust speech recognition. , 2015, , .		14
79	Local trajectory based speech enhancement for robust speech recognition with deep neural network. , 2015, , .		0
80	Deep feature for text-dependent speaker verification. Speech Communication, 2015, 73, 1-13.	2.8	124
81	An investigation on DNN-derived bottleneck features for GMM-HMM based robust speech recognition. , 2015, , .		6
82	Automatic model redundancy reduction for fast back-propagation for deep neural networks in speech recognition. , 2015, , .		6
83	Constrained Markov Bayesian Polynomial for Efficient Dialogue State Tracking. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23, 2177-2188.	5.8	13
84	Recurrent Polynomial Network for Dialogue State Tracking with Mismatched Semantic Parsers. , 2015, , .		3
85	Acoustic emotion recognition using deep neural network. , 2014, , .		14
86	An investigation of implementation and performance analysis of DNN based speech synthesis system. , 2014, , .		9
87	A generalized rule based tracker for dialogue state tracking. , 2014, , .		29
88	Evaluating vad for automatic speech recognition. , 2014, , .		7
89	Second order vector taylor series based robust speech recognition. , 2014, , .		3
90	Stochastic data sweeping for fast DNN training. , 2014, , .		3

#	ARTICLE	IF	CITATIONS
91	Speaker verification with deep features. , 2014, , .		7
92	Reshaping deep neural network for fast decoding by node-pruning. , 2014, , .		82
93	The SJTU System for Dialog State Tracking Challenge 2. , 2014, , .		36
94	Combination of data borrowing strategies for low-resource LVCSR. , 2013, , .		9
95	Review of F0 modelling and generation in HMM based speech synthesis. , 2012, , .		3
96	Continuous F0 Modeling for HMM Based Statistical Parametric Speech Synthesis. IEEE Transactions on Audio Speech and Language Processing, 2011, 19, 1071-1079.	3.2	86
97	Context adaptive training with factorized decision trees for HMM-based statistical parametric speech synthesis. Speech Communication, 2011, 53, 914-923.	2.8	29
98	On-line policy optimisation of spoken dialogue systems via live interaction with human subjects. , 2011, , .		36
99	Structured precision modelling with Cholesky Basis Superposition for speech recognition. , 2011, , .		0
100	Unsupervised training and directed manual transcription for LVCSR. Speech Communication, 2010, 52, 652-663.	2.8	47
101	The Hidden Information State model: A practical framework for POMDP-based spoken dialogue management. Computer Speech and Language, 2010, 24, 150-174.	4.3	321
102	Unsupervised Adaptation With Discriminative Mapping Transforms. IEEE Transactions on Audio Speech and Language Processing, 2009, 17, 714-723.	3.2	14
103	Bayesian Adaptive Inference and Adaptive Training. IEEE Transactions on Audio Speech and Language Processing, 2007, 15, 1932-1943.	3.2	16
104	Discriminative cluster adaptive training. IEEE Transactions on Audio Speech and Language Processing, 2006, 14, 1694-1703.	3.2	17
105	Karyotyping of comparative genomic hybridization human metaphases using kernel nearest-neighbor algorithm. Cytometry, 2002, 48, 202-208.	1.8	7
106	Kernel Nearest-Neighbor Algorithm. Neural Processing Letters, 2002, 15, 147-156.	3.2	104
107	How to optimize OCT image. Optics Express, 2001, 9, 24.	3.4	18
108	Rich Prosody Diversity Modelling with Phone-Level Mixture Density Network. , 0, , .		7

#	ARTICLE	IF	CITATIONS
109	What Does the Speaker Embedding Encode?. , 0, , .		27
110	Angular Softmax for Short-Duration Text-independent Speaker Verification. , 0, , .		60
111	Canonical state models for automatic speech recognition. , 0, , .		18
112	Voice Activity Detection in the Wild via Weakly Supervised Sound Event Detection. , 0, , .		6