

Akinori Ito

List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/6623585/publications.pdf>

Version: 2024-02-01

140
papers

448
citations

1040056

9
h-index

1199594

12
g-index

148
all docs

148
docs citations

148
times ranked

200
citing authors

#	ARTICLE	IF	CITATIONS
1	A Light-weight Hand-waving Gesture Recognition Method Using Kinect V2 and Frequency Analysis. , 2021, , .		0
2	SMOC corpus: A large-scale Japanese spontaneous multimodal one-on-one chat-talk corpus for dialog systems. Acoustical Science and Technology, 2021, 42, 210-213.	0.5	2
3	Development of a Mobile Robot That Plays Tag with Touch-and-Away Behavior Using a Laser Range Finder. Applied Sciences (Switzerland), 2021, 11, 7522.	2.5	3
4	Study on the Background Music Cancellation System for Speech Privacy. , 2021, , .		0
5	Analysis of Feature Extraction by Convolutional Neural Network for Speech Emotion Recognition. , 2021, , .		3
6	Automatic assessment of English proficiency for Japanese learners without reference sentences based on deep neural network acoustic models. Speech Communication, 2020, 116, 86-97.	2.8	14
7	A Symbol-level Melody Completion Based on a Convolutional Neural Network with Generative Adversarial Learning. Journal of Information Processing, 2020, 28, 248-257.	0.4	1
8	Evaluation of Person Tracking Methods for Human-Robot Physical Play. , 2020, , .		3
9	Language modeling in speech recognition for grammatical error detection based on neural machine translation. Acoustical Science and Technology, 2020, 41, 788-791.	0.5	5
10	Human-machine metacommunication towards development of a human-like agent: A short review. Acoustical Science and Technology, 2020, 41, 166-169.	0.5	0
11	Spoken Term Detection Based on Acoustic Models Trained in Multiple Languages for Zero-Resource Language. , 2020, , .		0
12	LJSing: Large-Scale Singing Voice Corpus of Single Japanese Singer. , 2020, , .		0
13	Improving Pronunciation Clarity of Dysarthric Speech Using CycleGAN with Multiple Speakers. , 2020, , .		0
14	Analysis and Estimation of Sentence Speakability for English Pronunciation Evaluation. , 2020, , .		1
15	Successive Japanese Lyrics Generation Based on Encoder-Decoder Model. , 2020, , .		1
16	Filler Prediction Based on Bidirectional LSTM for Generation of Natural Response of Spoken Dialog. , 2020, , .		1
17	A Study on Minimum Spectral Error Analysis of Speech. , 2020, , .		0
18	Integration of Accent Sandhi and Prosodic Features Estimation for Japanese Text-to-Speech Synthesis. , 2020, , .		0

#	ARTICLE	IF	CITATIONS
19	Incremental Response Generation Using Prefix-to-Prefix Model for Dialogue System. , 2020, , .		0
20	Realization of a Robot System That Plays "Darumasan-Ga-Koronda" Game with Humans. Robotics, 2019, 8, 55.	3.5	4
21	Improving human scoring of prosody using parametric speech synthesis. Speech Communication, 2019, 111, 14-21.	2.8	6
22	A Pedestrian Avoidance Method Considering Personal Space for a Guide Robot. Robotics, 2019, 8, 97.	3.5	3
23	Multi-condition training for noise-robust speech emotion recognition. Acoustical Science and Technology, 2019, 40, 406-409.	0.5	2
24	A Study on a Spoken Dialogue System with Cooperative Emotional Speech Synthesis Using Acoustic and Linguistic Information. Smart Innovation, Systems and Technologies, 2019, , 101-108.	0.6	1
25	Comparison of Speech Recognition Performance Between Kaldi and Google Cloud Speech API. Smart Innovation, Systems and Technologies, 2019, , 109-115.	0.6	7
26	Latent Words Recurrent Neural Network Language Models for Automatic Speech Recognition. IEICE Transactions on Information and Systems, 2019, E102.D, 2557-2567.	0.7	7
27	Melody Completion Based on Convolutional Neural Networks and Generative Adversarial Learning. Smart Innovation, Systems and Technologies, 2019, , 116-123.	0.6	0
28	Data Collection and Analysis for Automatically Generating Record of Human Behaviors by Environmental Sound Recognition. Smart Innovation, Systems and Technologies, 2019, , 149-156.	0.6	0
29	DNN-Based Talking Movie Generation with Face Direction Consideration. Smart Innovation, Systems and Technologies, 2019, , 157-164.	0.6	0
30	Muting Machine Speech Using Audio Watermarking. Smart Innovation, Systems and Technologies, 2019, , 74-81.	0.6	1
31	Leveraging a Small Corpus by Different Frame Shifts for Training of a Speech Recognizer. Smart Innovation, Systems and Technologies, 2019, , 82-89.	0.6	0
32	Improvement of Accent Sandhi Rules Based on Japanese Accent Dictionaries. Smart Innovation, Systems and Technologies, 2019, , 140-148.	0.6	0
33	Evaluation of English Speech Recognition for Japanese Learners Using DNN-Based Acoustic Models. Smart Innovation, Systems and Technologies, 2019, , 93-100.	0.6	2
34	Analysis of preferred speaking rate and pause in spoken easy Japanese for non-native listeners. Acoustical Science and Technology, 2018, 39, 92-100.	0.5	3
35	Domain Adaptation Based on Mixture of Latent Words Language Models for Automatic Speech Recognition. IEICE Transactions on Information and Systems, 2018, E101.D, 1581-1590.	0.7	4
36	Spoken Term Detection of Zero-Resource Language using Machine Learning. , 2018, , .		2

#	ARTICLE	IF	CITATIONS
37	Multiple player detection and tracking method using a laser range finder for a robot that plays with human. ROBOMECH Journal, 2018, 5, .	1.6	9
38	An Analysis of the Effect of Emotional Speech Synthesis on Non-Task-Oriented Dialogue System. , 2018, , .		9
39	A Study on 2D Photo-Realistic Facial Animation Generation Using 3D Facial Feature Points and Deep Neural Networks. Smart Innovation, Systems and Technologies, 2018, , 112-118.	0.6	1
40	Response Selection of Interview-Based Dialog System Using User Focus and Semantic Orientation. Smart Innovation, Systems and Technologies, 2018, , 84-90.	0.6	0
41	Dialog-Based Interactive Movie Recommendation: Comparison of Dialog Strategies. Smart Innovation, Systems and Technologies, 2018, , 77-83.	0.6	0
42	Voice Conversion from Arbitrary Speakers Based on Deep Neural Networks with Adversarial Learning. Smart Innovation, Systems and Technologies, 2018, , 97-103.	0.6	1
43	Evaluation of Nonlinear Tempo Modification Methods Based on Sinusoidal Modeling. Smart Innovation, Systems and Technologies, 2018, , 104-111.	0.6	0
44	Development and Evaluation of Julius-Compatible Interface for Kaldi ASR. Smart Innovation, Systems and Technologies, 2018, , 91-96.	0.6	1
45	Detection of Singing Mistakes from Singing Voice. Smart Innovation, Systems and Technologies, 2018, , 130-136.	0.6	0
46	Improving User Impression in Spoken Dialog System with Gradual Speech Form Control. , 2018, , .		5
47	Effect of Mutual Self-Disclosure in Spoken Dialog System on User Impression. , 2018, , .		3
48	Cluster-based approach to discriminate the user's state whether a user is embarrassed or thinking to an answer to a prompt. Journal on Multimodal User Interfaces, 2017, 11, 185-196.	2.9	1
49	Enhancement of person detection and tracking for a robot that plays with human. , 2017, , .		2
50	A Crowd Avoidance Method Using Circular Avoidance Path for Robust Person Following. Journal of Robotics, 2017, 2017, 1-10.	0.9	2
51	Analysis of efficient multimodal features for estimating user's willingness to talk: Comparison of human-machine and human-human dialog. , 2017, , .		6
52	Synthesis of Photo-Realistic Facial Animation from Text Based on HMM and DNN with Animation Unit. Smart Innovation, Systems and Technologies, 2017, , 29-36.	0.6	3
53	HMM-Based Photo-Realistic Talking Face Synthesis Using Facial Expression Parameter Mapping with Deep Neural Networks. Journal of Computer and Communications, 2017, 05, 50-65.	0.9	4
54	Collection of Example Sentences for Non-task-Oriented Dialog Using a Spoken Dialog System and Comparison with Hand-Crafted DB. Communications in Computer and Information Science, 2017, , 458-464.	0.5	0

#	ARTICLE	IF	CITATIONS
55	Investigation of Combining Various Major Language Model Technologies including Data Expansion and Adaptation. IEICE Transactions on Information and Systems, 2016, E99.D, 2452-2461.	0.7	4
56	Recognition of sounds using square cauchy mixture distribution. , 2016, , .		2
57	Construction and analysis of phonetically and prosodically balanced emotional speech database. , 2016, , .		22
58	A precise evaluation method of prosodic quality of non-native speakers using average voice and prosody substitution. , 2016, , .		1
59	Multiple description vector quantizer design based on redundant representation of central code. , 2016, , .		1
60	Tempo Modification of Mixed Music Signal by Nonlinear Time Scaling and Sinusoidal Modeling. , 2015, , .		1
61	Development of a mobile robot moving on a handrail —Control for preceding a person keeping a distance. , 2015, , .		2
62	Conversion of Speaker's Face Image Using PCA and Animation Unit for Video Chatting. , 2015, , .		4
63	Playing with a Robot: Realization of "Red Light, Green Light" Using a Laser Range Finder. , 2015, , .		9
64	Hierarchical Latent Words Language Models for Robust Modeling to Out-Of Domain Tasks. , 2015, , .		1
65	On Appropriateness and Estimation of the Emotion of Synthesized Response Speech in a Spoken Dialogue System. Communications in Computer and Information Science, 2015, , 747-752.	0.5	3
66	Speech recognition in a home environment using parallel decoding with GMM-based noise modeling. , 2014, , .		0
67	A study on the effect of speech rate on perception of spoken easy Japanese using speech synthesis. , 2014, , .		2
68	Subjective evaluation of packet loss recovery techniques for voice over IP. , 2014, , .		2
69	Assessing the Intended Enthusiasm of Singing Voice Using Energy Variance. , 2014, , .		0
70	Manipulation of vocal signal in mixed music signal using side information of F0 and backing spectrum. , 2014, , .		0
71	Analysis of English Pronunciation of Singing Voices Sung by Japanese Speakers. , 2014, , .		1
72	Automatic evaluation of singing enthusiasm for karaoke. Computer Speech and Language, 2014, 28, 501-517.	4.3	10

#	ARTICLE	IF	CITATIONS
73	Robot: Have I done something wrong? — Analysis of prosodic features of speech commands under the robot's unintended behavior. , 2014, , .		1
74	Teaching a robot where objects are: Specification of object location using human following and human orientation estimation. , 2014, , .		3
75	Controlling Switching Pause Using an AR Agent for Interactive CALL System. Communications in Computer and Information Science, 2014, , 588-593.	0.5	1
76	User Modeling by Using Bag-of-Behaviors for Building a Dialog System Sensitive to the Interlocutor's Internal State. , 2014, , .		7
77	Multi-modal Voice Activity Detection by Embedding Image Features into Speech Signal. , 2013, , .		0
78	ASAHI: OK for failure A robot for supporting daily life, equipped with a robot avatar. , 2013, , .		6
79	A Packet Loss Recovery of G.729 Speech Using Discriminative Model and N-Gram. , 2013, , .		2
80	Acoustic Features and Auditory Impressions of Death Growl and Screaming Voice. , 2013, , .		2
81	Speech recognition under noisy environments using multiple microphones based on asynchronous and intermittent measurements. , 2013, , .		0
82	Evaluation of Sinusoidal Modeling for Polyphonic Music Signal. , 2013, , .		5
83	Estimation of User’s State during a Dialog Turn with Sequential Multi-modal Features. Communications in Computer and Information Science, 2013, , 572-576.	0.5	2
84	Foreword to the special issue on "the speech communication and its related technologies". Acoustical Science and Technology, 2013, 34, 63-63.	0.5	0
85	Estimation of User's Internal State before the User's First Utterance Using Acoustic Features and Face Orientation. , 2012, , .		1
86	Model Shrinkage for Discriminative Language Models. IEICE Transactions on Information and Systems, 2012, E95.D, 1465-1474.	0.7	0
87	Automatic assessment of easiness of Japanese for writing aid of “Easy Japanese”. , 2012, , .		2
88	Effect of Robot Height on Comfortableness of Spoken Dialog. , 2012, , .		4
89	Spoken document retrieval by discriminative modeling in a high dimensional feature space. , 2012, , .		2
90	Estimating a User's Internal State before the First Input Utterance. Advances in Human-Computer Interaction, 2012, 2012, 1-10.	2.8	4

#	ARTICLE	IF	CITATIONS
91	Round-Robin Duel Discriminative Language Models. IEEE Transactions on Audio Speech and Language Processing, 2012, 20, 1244-1255.	3.2	16
92	Robust Transmission of Audio Signals over the Internet: An Advanced Packet Loss Concealment for MP3-Based Audio Signals. Interdisciplinary Information Sciences, 2012, 18, 99-105.	0.4	1
93	On short essays carried in the Acoustical Science and Technology. Acoustical Science and Technology, 2012, 33, 72.	0.5	0
94	Utterance Classification for Combination of Multiple Simple Dialog Systems. , 2011, , .		1
95	Manipulating Vocal Signal in Mixed Music Sounds Using Small Amount of Side Information. , 2011, , .		1
96	Toward Human-Robot Interaction Design through Human-Human Interaction Experiment. Lecture Notes in Electrical Engineering, 2011, , 127-130.	0.4	1
97	Bit rate reduction of the MELP coder using Lempel-Ziv segment quantization. , 2011, , .		3
98	Round-robin duel discriminative language models in one-pass decoding with on-the-fly error correction. , 2011, , .		3
99	Speech Recognition under Multiple Noise Environment Based on Multi-Mixture HMM and Weight Optimization by the Aspect Model. IEICE Transactions on Information and Systems, 2010, E93-D, 2407-2416.	0.7	0
100	Information Hiding for G.711 Speech Based on Substitution of Least Significant Bits and Estimation of Tolerable Distortion. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2010, E93-A, 1279-1286.	0.3	9
101	Improved Reference Speaker Weighting Using Aspect Model. IEICE Transactions on Information and Systems, 2010, E93-D, 1927-1935.	0.7	1
102	A spoken dialog system based on automatically-generated example database. , 2010, , .		2
103	Aspect-model-based reference speaker weighting. , 2010, , .		5
104	A query-by-humming music information retrieval from audio signals based on multiple F0 candidates. , 2010, , .		1
105	Document expansion using relevant web documents for spoken document retrieval. , 2010, , .		2
106	Improvement of Packet Loss Concealment for MP3 Audio Based on Switching of Concealment Method and Estimation of MDCT Signs. , 2010, , .		0
107	Source-filter separation for nonstationary voiced speech based on sinusoidal representation. Acoustical Science and Technology, 2010, 31, 181-184.	0.5	1
108	EFFECT OF THE SIZE FACTOR ON PSYCHOLOGICAL THREAT OF A MOBILE ROBOT MOVING TOWARD HUMAN. KANSEI Engineering International, 2009, 8, 51-58.	0.2	9

#	ARTICLE	IF	CITATIONS
109	A Band Extension of G.711 Speech with Low Computational Cost for Data Hiding Application. , 2009, , .		1
110	Information hiding for G.711 speech based on substitution of least significant bits and estimation of tolerable distortion. , 2009, , .		7
111	Data Hiding is a Better Way for Transmitting Side Information for MP3 Bitstream. , 2009, , .		0
112	Detection of Abnormal Sound Using Multi-stage GMM for Surveillance Microphone. , 2009, , .		9
113	A speaker adaptation method for non-native speech using learnersâ€™ native utterances for computer-assisted language learning systems. Speech Communication, 2009, 51, 875-882.	2.8	15
114	Relevant document retrieval using a spoken document. , 2009, , .		3
115	Multiple description coding for wideband audio signal transmission. , 2009, , .		1
116	Novel Tonal Feature and Statistical User Modeling for Query-by-Humming. Journal of Information Processing, 2009, 17, 95-105.	0.4	2
117	Automatic Query Generation and Query Relevance Measurement for Unsupervised Language Model Adaptation of Speech Recognition. Eurasip Journal on Audio, Speech, and Music Processing, 2009, 2009, 1-12.	2.1	2
118	EVALUATION OF ROBOT-AVATAR-BASED USER-FAMILIARITY IMPROVEMENT FOR ELDERLY PEOPLE. KANSEI Engineering International, 2009, 8, 59-66.	0.2	7
119	Bit rate reduction of mixed excitation linear prediction coder by Lempel-Ziv segment quantization. Acoustical Science and Technology, 2009, 30, 136-138.	0.5	0
120	Automatic clustering of part-of-speech for vocabulary divided PLSA language model. , 2008, , .		3
121	Intonation evaluation of English utterances using synthesized speech for Computer-Assisted Language Learning. , 2008, , .		0
122	Are bigger robots scary? —The relationship between robot size and psychological threat—. , 2008, , .		12
123	Packet Loss Concealment for MDCT-Based Audio Codec Using Correlation-Based Side Information. , 2008, , .		2
124	An unsupervised language model adaptation based on keyword clustering and query availability estimation. , 2008, , .		2
125	Increasing Correlation using a Few Bits for Multiple Description Coding. , 2007, , .		1
126	A New Segment Quantization Using Lempelâ€™Ziv Algorithm and Its Application to Quantization of Line Spectral Frequencies. IEEE Transactions on Communications, 2007, 55, 661-664.	7.8	1

#	ARTICLE	IF	CITATIONS
127	Pronunciation error detection for computer-assisted language learning system based on error rule clustering using a decision tree. <i>Acoustical Science and Technology</i> , 2007, 28, 131-133.	0.5	9
128	Multiple Description Coding of an Audio Stream by Optimum Recovery Transform. , 2006, , .		0
129	Music Information Retrieval from a Singing Voice Using Lyrics and Melody Information. <i>Eurasip Journal on Advances in Signal Processing</i> , 2006, 2007, 1.	1.7	23
130	An effective music information retrieval method using three-dimensional continuous DP. <i>IEEE Transactions on Multimedia</i> , 2006, 8, 633-639.	7.2	11
131	A grammatical error detection method for dialogue-based CALL system. <i>Journal of Natural Language Processing</i> , 2005, 12, 137-156.	0.2	6
132	Fast optimization of language model weight and insertion penalty from n-best candidates. <i>Acoustical Science and Technology</i> , 2005, 26, 384-387.	0.5	1
133	Error Tolerant Melody Matching Method in Music Information Retrieval. <i>Lecture Notes in Computer Science</i> , 2004, , 212-227.	1.3	0
134	Construction and evaluation of language models based on stochastic context-free grammar for speech recognition. <i>Systems and Computers in Japan</i> , 2002, 33, 48-59.	0.2	2
135	Language modeling by stochastic dependency grammar for Japanese speech recognition. <i>Systems and Computers in Japan</i> , 2001, 32, 10-15.	0.2	3
136	Japanese Dictation Toolkit. 1997 version.. <i>Journal of the Acoustical Society of Japan (E)</i> , 1999, 20, 233-239.	0.1	23
137	Performance prediction of word recognition using the probability of word occurrence. <i>Electronics and Communications in Japan, Part III: Fundamental Electronic Science (English Translation of Denshi Tj ETQq1 1 00784314 rgBT /Over</i>		
138	Performance prediction of word recognition using the transition information between phonemes or between characters. <i>Systems and Computers in Japan</i> , 1994, 25, 72-81.	0.2	2
139	Analysis of spectral enhancement using global variance in HMM-based speech synthesis. , 0, , .		4
140	Analyzing Effect of Physical Expression on English Proficiency for Multimodal Computer-Assisted Language Learning. , 0, , .		2