Mark Hasegawa-Johnson

M.E. Van Valkenburg Professor of Electrical and Computer Engineering
Beckman Institute 2011, University of Illinois, Urbana, IL 61801
Phone: 217-333-0925, Fax: 217-244-8371, jhasegaw@illinois.edu

October 25, 2025

Education

- **Post-Doctoral Fellow**, 1997-9, University of California at Los Angeles. Adviser: Abeer Alwan, Electrical Engineering. NRSA Title: Factor Analysis of MRI-Derived Articulator Shapes
- **Joint Post-Doctoral Fellow**, 1996-7, Massachusetts Institute of Technology and University of California at Los Angeles. Advisers: Kenneth N. Stevens (MIT) and Abeer Alwan (UCLA).
- **Ph.D. Electrical Engineering and Computer Science**, August 1996, Massachusetts Institute of Technology. Adviser: Kenneth N. Stevens. Thesis: Formant and Burst Spectral Measurements with Quantitative Error Models for Speech Sound Classification
- **M.S. and B.S. Electrical Engineering and Computer Science**, June 1989, Massachusetts Institute of Technology. Adviser: Jae S. Lim. Thesis: Echo Cancellation in the GSM Cellular Network

Full-Time Appointments

- 2023-present: **M.E. Van Valkenburg Professor of Electrical and Computer Engineering**, **University of Illinois**, Urbana, IL, USA. **Full-Time Faculty**, Beckman Institute for Advanced Science and Technology. **Affiliated Professor**, Coordinated Science Lab. **Affiliated Professor**, Graduate Program in Informatics. **Affiliated Professor**, Department of Computer Science.
- 2020-2023: **William L. Everitt Faculty Scholar in Electrical and Computer Engineering**, University of Illinois
- 2011-2020: **Professor**, Electrical and Computer Engineering, University of Illinois
- 2005-2011: **Associate Professor**, Electrical and Computer Engineering, University of Illinois
- 1999-2005: **Assistant Professor**, Electrical and Computer Engineering, University of Illinois
- 1997-1999: **Post-Doctoral Fellow**, **University of California at Los Angeles**, USA
- 1996-1997: **Joint Post-Doctoral Fellow**, **Massachusetts Institute of Technology** and **University of California at Los Angeles**, USA
- 1991-1996: **Graduate Research Assistant**, **Massachusetts Institute of Technology**, Cambridge, MA, USA
- 1989-1990: **Engineer**, **Fujitsu Laboratories Limited**, Kawasaki, Japan
- 1988-1989: **Engineering Intern**, **Motorola Corporate Research**, Schaumburg, IL, USA

Visiting, Affiliate, and Part-Time Appointments

- 2016-present: **Professor**, **Kyoto College of Graduate Studies for Informatics**, Kyoto, Japan
- 2022: **Visiting Professor**, Kyoto University, Kyoto, Japan
- 2019: **Visiting Professor**, Universidad Politécnica de Madrid, Madrid, Spain
- 2014-2018: **Research Professor**, **Advanced Digital Sciences Center**, Singapore
- 2011-2015: **Visiting Professor**, Qatar University, Doha, Qatar
- 2007: **Visiting Professor**, Kyoto University, Kyoto, Japan

Society Memberships

1. **Fellow** of the International Speech Communication Association (ISCA), for contributions to knowledge-constrained signal generation, 2023 (Member 2000-present)
2. **Fellow** of the IEEE, for contributions to speech processing of under-resourced languages, 2020 (Student Member 1988-1996, Member 1996-2004, Senior Member 2004-2020)
3. **Fellow** of the Acoustical Society of America, for contributions to vocal tract and speech modeling, 2011 (Student Member 1990-1996, Member 1996-2011)
4. **Senior Member**, Association for Computing Machinery, 2009 (Member 1999-2009)

Editorships

1. **Editor-in-Chief Elect,** IEEE Transactions on Audio, Speech and Language, 2025-
2. **Deputy Editor,** IEEE Transactions on Audio, Speech and Language, 2023-2024
3. **Senior Area Editor,** IEEE Transactions on Audio, Speech and Language, 2018-2023
4. **Associate Editor,** J. Acoust. Soc. Am. (2009-2017), Laboratory Phonology (2009-2015), IEEE Trans. Audio, Speech, and Language (2006-2009), IEEE Signal Processing Letters (2002-2004)

Technical Conference and Workshop Offices Held

1. **Technical Program Chair,** IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2025
2. **UIUC Representative,** Midwest Speech and Language Days, 2009-2024
3. **Area Chair,** Interspeech 2017, 2019-20, 2023
4. **Technical Program Chair,** Interspeech 2022
5. **Workshop Co-Chair,** SPREI Speech Production Workshop, Urbana, IL, May 2011
6. **Conference General Chair,** Fifth International Conference on Speech Prosody, 2010
7. **Workshops Co-Chair,** HLT/NAACL 2009
8. **Technical Committee or Conference Reviewer:** AAAI; ACL; AISTATS; UIUC Allerton; APSIPA; IEEE ASRU; CBMI; EACL; EMNLP; ACL HLT; IEEE ICASSP; IEEE ICC; ICLR; ICML; ICPR; ICPPIT; ICPRAM; ISCA Interspeech; LabPhon; MCLC; NAACL; NeurIPS; ISCA SigML; ISCA/ACL SLPAT; ISCA SLaTE; ISCA SPASR; UIUC SPREI; IEEE SLT

Professional Society Offices Held

1. **Member of the Steering Committee,** IEEE Signal Processing Society Mentoring Experiences for Underrepresented Young Researchers, 2022-present
2. **Member, Diversity Committee,** International Speech Communication Association (ISCA), 2017-present
3. **Treasurer, ISCA,** 2013-2021
4. **Society Liaison Officer,** ISCA Special Interest Group on Machine Learning (SIGML), 2010-2020
5. **Secretary,** ISCA Speech Prosody Special Interest Group (SProSIG), 2010-2018
6. **Member, Speech and Language Technical Committee (SLTC)**, IEEE Signal Processing Society, 2011-2017

Service to Student Organizations

1. **Phi Beta Kappa:** University of Illinois chapter Executive Secretary (2006-2017, 2019-present), Scholarship Chair (2004-2006)
2. **Robocup,** University of Illinois advisor, 2020-2021
3. **Eta Kappa Nu** University of Illinois chapter advisor, 2004-7

Conference and Workshop Keynotes, Invited Talks, and Panels

1. **Panelist,** "Voice-based AI for Medicine: Practices for Ethical Interventions at Scale," panel at the American Academy of Arts and Sciences (AAAS), Boston, February 14, 2025
2. **Keynote,** "A Theory of Unsupervised Speech Recognition," IEEE Workshop on Spoken Language Technology (SLT), Macau, December 5, 2024
3. **Invited talk,** "Variability in the Speech Accessibility Project," Automatic Assessment of Parkinsonian Speech Workshop AAPS'2024, Cambridge, MA, November 21-22,
4. **Panelist,** "AI's Impact on Society," panel at the NSF Summit for AI Institute Leaders (SAIL), Pittsburgh, October 8, 2024
5. **Invited talk,** "Community-supported shared infrastructure in support of speech accessibility," Convention of the American Speech-Language Hearing Association (ASHA), Boston, November 2023
6. **Invited talk,** "Grapheme-to-Phoneme Transduction for Cross-Language ASR," Workshop on Statistical Language and Speech Processing (SLSP), 2021
7. **Invited talk,** "Multimodal Distant Supervision," NeurIPS workshop on Self-Supervised Learning for Speech and Audio Processing, December, 2020
8. **Keynote,** "Distant Supervision for Cross-Language Speech Adaptation." IEEE Workshop on Automatic Speech Recognition and Understanding, Singapore, December, 2019
9. **Invited talk,** "Multimedia AI and User Trust," AI Association of Korea Workshop on Artificial Intelligence and Blockchain, Seoul, November, 2019
10. **Keynote,** "Unwritten Languages as a Test Case for the Theory of Phonetic Universals," International Symposium on Chinese Spoken Language Processing, Taipei, November, 2018
11. **Invited lecture series,** "Prosody in Speech Technology," ISCA Winter School on Speech and Audio Processing, Chennai, December, 2016
12. **Panel Organizer,** VAC Consortium Working Group on Multimedia Analytics, Adelphi, MD, May 2011

Awards

1. **List of Teachers Rated "Excellent" by their Students**, University of Illinois at Urbana-Champaign, eleven times including every year 2019-2024.
2. **Outstanding Reviewer,** ICASSP 2023
3. **Top 3%** of ICASSP 2023 papers, for the paper "Dual-Path Cross-Modal Attention for better Audio-Visual Speech Extraction," Zhongweiyang Xu, Xulin Fan, and Mark Hasegawa-Johnson.
4. **Second Place,** Diagnosing COVID-19 from Acoustics (DiCOVA) challenge competition, ICASSP 2022
5. **Third Place,** Diagnosing COVID-19 from Acoustics (DiCOVA) challenge competition, Interspeech 2021
6. **William L. Everitt Faculty Scholar Award**, ECE Department, University of Illinois, 2020-2021
7. **Best Paper Award**, IEEE Transactions on Audio Speech and Language 2020, for Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, and Paris Smaragdis, "Joint Optimization of Masks and Deep Recurrent Neural Networks for Monaural Source Separation,"
8. **Best Reviewer**, Neural Information Processing Systems (NIPS), 2005 and 2020
9. **Best Student Paper**, for the Paper "Adapting ASR for Under-Resourced Languages Using Mismatched Transcriptions," Chunxi Liu, Preethi Jyothi, Hao Tang, Vimal Manohar, Rose Sloan, Tyler Kekona, Mark Hasegawa-Johnson, Sanjeev Khudanpur, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, March 2016

10. Starkey Grant for **Best Student Paper** in the AASP Area, for the Paper "Deep Learning for Monaural Speech Separation," Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, and Paris Smaragdis, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, May 4-9, 2014.
11. **Dean's Award for Excellence in Research**, University of Illinois College of Engineering, 2012
12. **Best Student Paper** for the paper "A Novel Gaussianized Vector Representation for Natural Scene Categorization," Xi Zhou, Xiaodan Zhuang, Hao Tang, Mark Hasegawa-Johnson, and Thomas Huang, *International Conference on Pattern Recognition (ICPR)*, Tampa, USA, Dec. 2008
13. **Third Place**, Star Challenge Multimedia Information Retrieval Competition, A*STAR 2008
14. **First Place,** Acoustic Event Recognition competition, 2007 CLEAR (Classification of Events, Activities and Relationships) Evaluation and Workshop (Stiefelhagen et al., "The CLEAR 2007 Evaluation," LNCS 4625:3-34, 2008)
15. **CINE Golden Eagle Award** for excellence in cinematic achievements, Non-Telecast - Independent Division: Fiction, Short: Producer, "Crystal's Sigh," 2006
16. **Outstanding Advisers List**, University of Illinois College of Engineering, April 2006
17. **Honorary Initiate**, Alpha Chapter of Eta Kappa Nu (Electrical and Computer Engineering Honor Society), 2003
18. **Frederick V. Hunt Post-Doctoral Research Fellow**, Acoustical Society of America, 1996
19. **Paul L. Fortescue Graduate Fellow**, IEEE, 1989
20. Undergraduate member of four honor societies: Eta Kappa Nu, Tau Beta Pi, Sigma Xi, and Phi Beta Kappa

Summer Intensive Research Programs

1. **Team Member,** Jelinek Speech and Language Technology (JSALT) Workshop WS17, "The Speaking Rosetta Stone - Discovering Grounded Linguistic Units for Languages without Orthography," Pittsburgh, June–August, 2017
2. **Team Leader,** Jelinek Speech and Language Technology (JSALT) Workshop WS15, "Probabilistic Transcription Using EEG and Crowdsourcing for Languages with No Native Language Transcribers," Seattle, June–August, 2015
3. **Team Member,** DARPA/NSF CLSP Summer Research Workshop, Articulatory-Feature Based Speech Recognition, Baltimore, MD, June–August, 2006
4. **Team Leader,** DARPA/NSF CLSP Summer Research Workshop, Landmark-Based Speech Recognition, Baltimore, MD, June–August, 2004

Course Director

1. **Multimedia Signal Processing (ECE 417)**, 2013-present
2. **Signal Processing (ECE 401)**, 2013-present
3. **Audio Engineering (ECE 403)**, 2001-2012

Ph.D. Students and Post-Doctoral Fellows

Advisees include eight post-doctoral researchers, thirty recipients of the Ph.D., and twenty-two recipients of the M.S. Alums include current faculty at Inha University, IIT Bombay, Purdue, UT Austin, and UT MD Anderson Biomedical Campus, and current research and technical staff members at Adobe, Amazon, Apple, The Aspen Group, A∗STAR, Credit Karma, DeepMind, EquiLibre Technologies, Google, Haven Technologies, IBM, Juniper Networks, KPMG, Meta, Microsoft, Novumind, NVIDIA, Otter.ai, Presto, Rambus, Samsung, ServiceNow, Shure, Tencent, Tesla, Unisound, and ZipRecruiter. A list is available at `https://speechtechnology.web.illinois.edu/people/`.

Research Grants

[1] Inha University, Ewha Women's University, and University of Illinois at Urbana-Champaign. Workshop on assistive and inclusive technology for digital accessibility. Funded Grant Number, 2 2025.

[2] Nancy McElwain, Mark Hasegawa-Johnson, and Bashima Islam. Automated assessment of infant sleep/wake states, physical activity, and household noise using a multimodal wearable device and deep learning models. Funded Grant Number 1 R01 DK138866-01, National Institute on Diabetes and Digestive and Kidney Diseases ($606,042), 2024–2029.

[3] Nancy McElwain, Mark Hasegawa-Johnson, Bashima Islam, Brandon Meline, Maria Davila, and Keri Heilman. Validation of a virtual still face procedure and deep learning algorithms to assess infant emotion regulation and infant-caregiver interactions in the wild. Funded Grant Number 1 R01 DA059422-01, National Institute on Drug Abuse ($1,280,875), 2024–2029.

[4] Venu Govindaraju, Jinjun Xiong, Srirangaraj Setlur, Pamela Hadley, Julie Kientz, David Feil-Seifer, Brian Graham, Megan-Brette Hamilton, Anil K Jain, Ye Jia, Jennifer Taps Richard, Tracy A Sawicki, Alison Hendricks, Ifeoma Nwogu, Letitia Thomas, Christine Wang, Wenyao Xu, Maneesh Agrawala, Diego Aguirre, Changyou Chen, Karthik Dantu, Mark Frank, Nick Haber, Mark Hasegawa-Johnson, Yun Huang, Heng Ji, Windi Krok, Hedda Meadan-Kaplansky, Carol Miller, Abbie Olszewski, Mari Ostendorf, Alexander Rush, Humphrey Shi, Hariharan Subramonyam, Nigel Ward, Junsong Yuan, and Shaofeng Zou. Ai institute for transforming education for children with speech and language processing challenges. Funded Grant Number 2229873, NSF DRL Division Of Research On Learning ($20,000,000), 2023–2028.

[5] Mark Hasegawa-Johnson. Unrestricted gift. Funded Grant Number 638245, Picsart Corporation ($100,000), 2022–2025.

[6] Mark Hasegawa-Johnson, Zsuzsanna Fagyal, Najim Dehak, Piotr Zelasko, and Laureano Moro-Velazquez. FAI: A new paradigm for the evaluation and training of inclusive automatic speech recognition. Funded Grant Number 2147350, NSF IIS Division of Information and Intelligent Systems ($500,000), 2022–2026.

[7] Mark Hasegawa-Johnson, Heejin Kim, Clarion Mendes, Meg Dickinson, Erik Hege, and Chris Zwilling. Speech accessibility project. Funded Grant Number 548308-392030, AI Accessibility Coalition (Amazon, Apple, Google, Meta and Microsoft), 2022–2025.

[8] Amit Juneja, Mark Hasegawa-Johnson, and James Mireles. Sttr phase ii proposal f2-15825 virtual reality visualization of complex and unstructured data andrelationships. Funded Grant Number F2-15825, USAF United States Air Force ($749,952), 2022–2023.

[9] Amit Juneja, Mark Hasegawa-Johnson, and James Mireles. Sttr virtual reality visualization of complex and unstructured data and relationships. Funded Grant Number FA864922P0047, USAF United States Air Force ($49,995), 2021–2022.

[10] Narendra Ahuja, Mark Hasegawa-Johnson, David Beiser, and David Chestek. Adding audio-visual cues to signs and symptoms for triaging suspected or diagnosed covid-19 patients. Funded Grant Number 23, C3.ai Digital Transformation Institute ($498,998), 2020–2022.

[11] Changdong Yoo and Mark Hasegawa-Johnson. Deep f-measure maximization for fairness in speech understanding. Funded Grant Number 507691, IITP Institute for Information & Communications Technology Promotion ($500,000), Korea, 2020–2025.

[12] Najim Dehak and Mark Hasegawa-Johnson. RI: Small: Collaborative research: Automatic creation of new phone inventories. Funded Grant Number 19-10319, NSF IIS Division of Information and Intelligent Systems ($259,765), 2019–2022.

[13] Nancy Chen and Mark Hasegawa-Johnson. Mismatched crowdsourcing for 80-language speech recognition. Funded Grant Number 36, Institute for Infocomm Research (I^2R), Agency for Science, Technology, Advancement and Research (ASTAR) ($399,999), Singapore, 2015–2017.

[14] Mark Hasegawa-Johnson. Languagenet: Transfer learning across a language similarity network. Funded Grant Number HR0011-15-2-0043, DARPA LO-REsource Languages for Emergent Incidents (LORELEI) ($408,325), 2015–2019.

[15] Mark Hasegawa-Johnson, Gautham Mysore, and Yang Zhang. Speech resynthesis for vocal style modification. Funded Grant Number 2015.m.uiuc, Adobe Research ($25,000), 2015–2017.

[16] Mark Hasegawa-Johnson, Lav Varshney, and Preethi Jyothi. EAGER: matching non-native transcribers to the distinctive features of the language transcribed. Funded Grant Number 1550145, NSF IIS Division of Information and Intelligent Systems ($150,000), 2015–2018.

[17] Mark Hasegawa-Johnson, Gregg Wilensky, and Xuesong Yang. Speech2vec speech-based semantic vectors. Funded Grant Number 2015.w.uiuc, Adobe Research ($25,000)), 2015–2016.

[18] Mark Hasegawa-Johnson, Hanady Mansour Ahmed, Eiman Mustafawi, and Haitham El-Bashir. The family as the unit of intervention for speech-generating augmentative/assistive communication. Funded Grant Number NPRP 7-766-1-140, QNRF Qatar National Research Fund ($858,827), 2014–2018.

[19] Daniel G Morrow, Mark Hasegawa-Johnson, Thomas Huang, and William Schuh. Collaborative patient portals: Computer-based agents and patients' understanding of numeric health information. Funded Grant Number R21-HS022948, AHRQ Agency for Healthcare Research and Quality ($299,602), 2014–2016.

[20] Mark Hasegawa-Johnson and Eiman Mustafawi. Multi-dialect phrase-based speech recognition and machine translation for Qatari broadcast tv. Funded Grant Number NPRP 09-410-1-069, QNRF Qatar National Research Fund ($734,266), 2010–2013.

[21] Marshall Poole, David Forsyth, Feniosky Pena-Mora, Mark Hasegawa-Johnson, Kenton McHenry, and Peter Bajcsy. CDI-Type II: collaborative research: Groupscope: Instrumenting research on interaction networks in complex social contexts. Funded Grant Number 0941268, NSF BCS Division Of Behavioral and Cognitive Sci ($1,697,482), 2010–2015.

[22] Richard Beraniuk, Volkan Cevher, Lydia Kavraki, Wotao Yin, John Benedetto, Rama Chellappa, Larry Davis, Tamer Basar, Mark Hasegawa-Johnson, Thomas Huang, Ronald Coifman, Lawrence Carin, and Stanley Osher. Opportunistic sensing for object and activity recognition from multi-modal, multi-platform data. Funded Grant Number W911NF-09-1-0383, ARO MURI, 2009–2014.

[23] Mark Hasegawa-Johnson, Camille Goudeseune, Hank Kaczmarski, and Thomas Huang. FODAVA-partner: Visualizing audio for anomaly detection. Funded Grant Number 08-07329, NSF CCF Division of Computing and Communication Foundations ($793,857), 2008–2013.

[24] Mark Hasegawa-Johnson, Thomas Huang, and Dirk Bernhardt-Walther. RI medium: Audio diarization - towards comprehensive description of audio events. Funded Grant Number 08-03219, NSF IIS Division of Information and Intelligent Systems ($249,864), 2008–2010.

[25] Jennifer Cole and Mark Hasegawa-Johnson. RI-collaborative research: Landmark-based robust speech recognition using prosody-guided models of speech variability. Funded Grant Number 07-03624, NSF IIS Division of Information and Intelligent Systems ($541,320), 2007–2012.

[26] Mark Hasegawa-Johnson, Adrienne Perlman, Thomas S Huang, Jon Gunderson, Ken Watkin, and Heejin Kim. Audiovisual description and recognition of audible and visible dysarthric phonology. Funded Grant Number 5R21 DC008090, NIH NIDCD ($388,651)), 2006–2009.

[27] Richard Sproat, J. Kathryn Bock, Brian Ross, Mark Hasegawa-Johnson, and Chi-Lin Shih. DHB: an interdisciplinary study of the dynamics of second language fluency. Funded Grant Number 06-23805, NSF IIS Division of Information and Intelligent Systems ($710,781), 2006–2010.

[28] Mark Hasegawa-Johnson, Adrienne Perlman, Thomas Huan, and Jon Gunderson. Audiovisual distinctive-feature-based recognition of dysarthric speech. Funded Grant Number 05-34106, NSF IIS Division of Information and Intelligent Systems ($668,575), 2005–2010.

[29] Thomas S Huang and Mark Hasegawa-Johnson. Audiovisual emotional speech avatar. Funded Grant Number RPS 31, Motorola Communications Center ($100,000), 2005–2007.

[30] Mark Hasegawa-Johnson, Jennifer Cole, and Chi-Lin Shih. Prosodic, intonational, and voice quality correlates of disfluency. Funded Grant Number 04-14117, NSF IIS Division of Information and Intelligent Systems ($526,973), 2004–2007.

[31] Mark Hasegawa-Johnson. Acoustic features for phoneme recognition. Funded Grant Number 2002.1.4, Phonetact Incorporated ($35,000), 2002.

[32] Mark Hasegawa-Johnson. Career: Landmark-based speech recognition in music and speech backgrounds. Funded Grant Number 01-32900, NSF IIS Division of Information and Intelligent Systems ($418,766), 2002–2007.

[33] Mark Hasegawa-Johnson and Thomas S Huang. Audiovisual speech recognition: Data collection and feature extraction in automotive environment. Funded Grant Number RPS 19, Motorola Communications Center ($100,000), 2002–2005.

[34] Weimo Zhu. Development and validation of an e-diary system for assessing physical activity and travel behaviors. Funded Grant Number 47334, Robert Wood Johnson Foundation ($100,000), 2002–2004.

[35] Mark Hasegawa-Johnson. Factor analysis of mri-derived articulator shapes. Funded Grant Number 1 F32 DC000323, NIH NIDCD ($145,026), 1997–1999.

Patents

[1] Mark Hasegawa-Johnson. Probabilistic segmental model for doppler ultrasound heart rate monitoring. United States Patent Number 8727991, May 2014.

[2] Tomohiko Taniguchi and Mark Johnson. Speech coding and decoding system. United States Patent Number 5799131, August 1998.

[3] Tomohiko Taniguchi, Mark Johnson, Yasuji Ohta, Hideki Kurihara, Yoshinori Tanaka, and Yoshihito Sakai. Speech coding system having codebook storing differential vectors between each two adjoining code vectors. United States Patent Number 5323486, June 1994.

[4] Tomohiko Taniguchi and Mark Johnson. Speech coding system. United States Patent Number 5245662, September 1993.

[5] Tomohiko Taniguchi, Mark Johnson, Hideki Kurihara, Yoshinori Tanaka, and Yasuji Ohta. Speech coding and decoding system. United States Patent Number 5199076, March 1993.

Journal Articles and JASA Abstracts

[1] Stefanie Shattuck-Hufnagel, Maria-Gabriella Di Benedetto, Abeer Alwan, and Mark Hasegawa-Johnson. Introduction to special issue on acoustic cue-based perception and production of speech by humans and machines. *J. Acoust. Soc. Am.*, 157(3):2102–2105, 3 2025.

[2] John Harvill, Moitreya Chatterjee, Shaveta Khosla, Mustafa Alam, Narendra Ahuja, Mark Hasegawa-Johnson, David Chestek, and David Beiser. Multimodal respiratory rate estimation from audio and video in emergency department patients. *IEEE Journal of Translational Engineering in Health and Medicine*, 2024.

[3] Mark Hasegawa-Johnson, Xiuwen Zheng, Heejin Kim, Clarion Mendes, Meg Dickinson, Erik Hege, Chris Zwilling, Marie Moore Channell, Laura Mattie, Heather Hodges, Lorraine Ramig, Mary Bellard, Mike Shebanek, Leda Sarı, Kaustubh Kalgaonkar, David Frerichs, Jeffrey P. Bigham, Leah Findlater, Colin Lea, Sarah Herrlinger, Peter Korn, Shadi Abou-Zahra, Rus Heywood, Katrin Tomanek, and Bob MacDonald. Community-supported shared infrastructure in support of speech accessibility. *Journal of Speech Language and Hearing Research*, 67(10):4162–4175, 2024.

[4] Bashima Islam, Nancy L McElwain, Jialu Li, Maria Davila, Yannan Hu, Kexin Hu, Jordan M Bodway, Ashutosh M Dhekne, Romit Roy Choudhury, and Mark Hasegawa-Johnson. Preliminary technical validation of littlebeats™: A multimodal sensing platform to capture cardiac physiology, motion, and vocalizations. *Sensors*, 24(3), 1 2024.

[5] Nancy McElwain, Bashima Islam, Meghan Fisher, Camille Nebeker, Jordan Marie Bodway, and Mark Hasegawa-Johnson. Evaluating users' experiences of a child multimodal wearable device: A mixed methods approach. *JMIR Human Factors*, 11:e49316, 2024.

[6] Mahir Morshed and Mark Hasegawa-Johnson. Using articulatory feature detectors in progressive networks for multilingual low-resource phone recognition. *J. Acoust. Soc. Am.*, 156(5):3411–3421, 11 2024.

[7] Abhayjeet Singh, Amala Nagireddi, Anjali Jayakumar, Deekshitha G, Jesuraja Bandekar, Roopa R, Sandhya Badiger, Sathvik Udupa, Saurabh Kumar, Prasanta Kumar Ghosh, Hema A Murthy, Heiga Zen, Pranaw Kumar, Kamal Kant, Amol Bole, Bira Chandra Singh, Keiichi Tokuda, Mark Hasegawa-Johnson, and Philipp Olbrich. Lightweight, multi-speaker, multi-lingual indic text-to-speech. *IEEE Open Journal of Signal Processing*, 5:790–798, 3 2024.

[8] Renato Azevedo, Rocio Garcia-Retamero, Mark Hasegawa-Johnson, Kuangxiao Gu, and Danel G. Morrow. The influence of memory for and affective response to health messages on self-care behavioral intentions. *Human Factors in Healthcare*, 4:100058, 12 2023.

[9] Eugene Cox, Mark Hasegawa-Johnson, Suma Bhat, Mukhil Umashankar, H Chad Lane, and Daniel Morrow. The importance of diverse user goals when designing an automated COVID risk counselor. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, 12(1):35–39, 3 2023.

[10] Heting Gao, Junrui Ni, Yang Zhang, Kaizhi Qian, Shiyu Chang, and Mark Hasegawa-Johnson. Domain generalization for language-independent automatic speech recognition. *Frontiers in Artificial Intelligence*, 5:806274, 2022.

[11] Heting Gao, Xiaoxuan Wang, Sunghun Kang, Rusty Mina, Dias Issa, John Harvill, Leda Sarı, Mark Hasegawa-Johnson, and Chang D. Yoo. Seamless equal accuracy ratio for inclusive ctc speech recognition. *Speech Communication*, 136:76–83, 2022.

[12] Jialu Li and Mark Hasegawa-Johnson. Autosegmental neural nets 2.0: An extensive study of training synchronous and asynchronous phones and tones for under-resourced tonal languages. *IEEE Transactions on Audio, Speech and Language*, 30:1918–1926, 5 2022.

[13] Jialu Li, Mark Hasegawa-Johnson, and Nancy L McElwain. Visualizations of complex sequences of family-infant vocalizations using bag-of-audio-words approach based on wav2vec 2.0 features. *arXiv preprint arXiv:2203.15183*, 2022.

[14] Piotr Zelasko, Siyuan Feng, Laureano Moro-Velazquez, Ali Abavisani, Saurabchand Bhati, Odette Scharenborg, Mark Hasegawa-Johnson, and Najim Dehak. Discovering phonetic inventories with crosslingual automatic speech recognition. *Computer Speech and Language*, 74:101358:1–54, 7 2022.

[15] Jialu Li, Mark Hasegawa-Johnson, and Nancy McElwain. Analysis of acoustic and voice quality features for the classification of infant and mother vocalizations. *Speech Communication*, 133:41–61, 2021.

[16] Leda Sarı, Mark Hasegawa-Johnson, and Chang D. Yoo. Counterfactually fair automatic speech recognition. *IEEE Transactions on Audio, Speech, and Language*, 29:3515–3525, 2021.

[17] Odette Scharenborg and Mark Hasegawa-Johnson. Position paper: Brain signal-based dialogue systems. *Lecture Notes in Computer Science*, 714, 3 2021.

[18] Mark Hasegawa-Johnson, Leanne Rolston, Camille Goudeseune, Gina-Anne Levow, and Katrin Kirchhoff. Grapheme-to-phoneme transduction for cross-language asr. *Lecture Notes in Computer Science*, 12379:3–19, 2020.

[19] Liming Wang and Mark Hasegawa-Johnson. Multimodal word discovery and retrieval with spoken descriptions and visual concepts. *IEEE Transactions on Audio, Speech and Language*, 28:1560–1573, 2020.

[20] Junrui Ni, Mark Hasegawa-Johnson, and Odette Scharenborg. The time-course of phoneme category adaptation in deep neural networks. *Lecture Notes in Artificial Intelligence*, 11816:3–18, October 2019.

[21] Daniel Morrow, Renato Azevedo, Leitão Ferreira, Rocio Garcia-Retamero, Mark Hasegawa-Johnson, Thomas Huang, William Schuh, Kuangxiao Gu, and Yang Zhang. Contextualizing numeric clinical test results for gist comprehension: Implications for EHR patient portals. *Journal of Experimental Psychology: Applied*, 25(1):41–61, 2019.

[22] Kaizhi Qian, Yang Zhang, Shiyu Chang, Xuesong Yang, and Mark Hasegawa-Johnson. Autovc: Zero-shot voice style transfer with only autoencoder loss. *Proceedings of Machine Learning Research*, 97:5210–5219, 2019.

[23] Van Hai Do, Nancy F. Chen, Boon Pang Lim, and Mark Hasegawa-Johnson. Multitask learning for phone recognition of underresourced languages using mismatched transcription. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 26(3):501–514, March 2018.

[24] Di He, Boon Pang Lim, Xuesong Yang, Mark Hasegawa-Johnson, and Deming Chen. Acoustic landmarks contain more information about the phone string than other frames for automatic speech recognition with deep neural network acoustic model. *Journal of the Acoustical Society of America*, 143(6):3207–3219, 2018.

[25] Mark Hasegawa-Johnson, Preethi Jyothi, Daniel McCloy, Majid Mirbagheri, Giovanni di Liberto, Amit Das, Bradley Ekin, Chunxi Liu, Vimal Manohar, Hao Tang, Edmund C. Lalor, Nancy Chen, Paul Hager, Tyler Kekona, Rose Sloan, and Adrian KC Lee. ASR for under-resourced languages from probabilistic transcription. *IEEE/ACM Trans. Audio, Speech and Language*, 25(1):46–59, 2017.

[26] Daniel Morrow, Mark Hasegawa-Johnson, Thomas Huang, William Schuh, Renato Azevedo, Kuangxiao Gu, Yang Zhang, Bidisha Roy, and Rocio Garcia-Retamero. A multidisciplinary approach to designing and evaluating electronic medical record portal messages that support patient self-care. *Journal of Biomedical Informatics*, 69:63–74, 5 2017.

[27] Wenda Chen, Mark Hasegawa-Johnson, and Nancy F. Chen. Mismatched crowdsourcing based language perception for under-resourced languages. *Procedia Computer Science*, 81:23–29, 2016.

[28] Karen Livescu, Frank Rudzicz, Eric Fosler-Lussier, Mark Hasegawa-Johnson, and Jeff Bilmes. Speech production in speech technologies: Introduction to the csl special issue. *Computer Speech and Language*, 36:165–172, 2016.

[29] Tom Le Paine, Pooya Khorrami, Shiyu Chang, Yang Zhang, Prajit Ramachandran, Mark A Hasegawa-Johnson, and Thomas S Huang. Fast wavenet generation algorithm. *arXiv preprint arXiv:1611.09482*, 2016.

[30] Mark Hasegawa-Johnson, Jennifer Cole, Preethi Jyothi, and Lav Varshney. Models of dataset size, question design, and cross-language speech perception for speech crowdsourcing applications. *Journal of Laboratory Phonology*, 6(3-4):381–431, 2015.

[31] Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, and Paris Smaragdis. Joint optimization of masks and deep recurrent neural networks for monaural source separation. *IEEE Trans. Audio, Speech and Language Processing*, 23(12):2136–2147, 2015.

[32] Harsh Vardhan Sharma and Mark Hasegawa-Johnson. Acoustic model adaptation using in-domain background models for dysarthric speech recognition. *Computer Speech and Language*, 27(6):1147–1162, September 2013.

[33] Robert Mertens, Po-Sen Huang, Luke Gottlieb, Gerald Friedland, Ajay Divakaran, and Mark Hasegawa-Johnson. On the application of speaker diarization to audio indexing of non-speech and mixed non-speech/speech video soundtracks. *International Journal of Multimedia Data Engineering and Management (IJDEM)*, 3(3):1–19, April 2013.

[34] Panying Rong, Torrey Loucks, Heejin Kim, and Mark Hasegawa-Johnson. Relationship between kinematics, f2 slope and speech intelligibility in dysarthria due to cerebral palsy. *Clinical Linguistics and Phonetics*, 26(9):806–822, September 2012.

[35] Hosung Nam, Vikramjit Mitra, Mark Tiede, Mark Hasegawa-Johnson, Carol Espy-Wilson, Elliot Saltzman, and Louis Goldstein. A procedure for estimating gestural scores from speech acoustics. *J. Acoustical Society of America*, 132(6):3980–3989, 2012.

[36] Hao Tang, Stephen Chu, Mark Hasegawa-Johnson, and Thomas Huang. Partially supervised speaker clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):959–971, 2012.

[37] Heejin Kim, Mark Hasegawa-Johnson, and Adrienne Perlman. Temporal and spectral characteristics of fricatives in dysarthria. *Journal of the Acoustical Society of America*, 130:2446, 2011.

[38] İ. Yücel Ozbek, Mark Hasegawa-Johnson, and Mübeccel Demirekler. Estimation of articulatory trajectories based on gaussian mixture model (gmm) with audio-visual information fusion and dynamic kalman smoothing. *IEEE Transactions on Audio, Speech, and Language*, 19(5):1180–1195, 2011.

[39] İ. Yücel Ozbek, Mark Hasegawa-Johnson, and Mübeccel Demirekler. On improving dynamic state space approaches to articulatory inversion with map based parameter estimation. *IEEE Transactions on Audio, Speech, and Language*, 20(1):67–81, 2011.

[40] Lae-Hoon Kim, Kyungtae Kim, and Mark Hasegawa-Johnson. Speech enhancement beyond minimum mean squared error with perceptual noise shaping. *J. Acoust. Soc. Am.*, 127(3):1817, 2010.

[41] Bryce E Lobdell, Jont B Allen, and Mark A Hasegawa-Johnson. Intelligibility predictors and neural representation of speech. *Speech Communication*, 2010.

[42] Thomas S. Huang, Mark A. Hasegawa-Johnson, Stephen M. Chu, Zhihong Zeng, and Hao Tang. Sensitive talking heads. *IEEE Signal Processing Magazine*, 26(4):67–72, July 2009.

[43] Mark Hasegawa-Johnson, Xiaodan Zhuang, Xi Zhou, Camille Goudeseune, and Thomas S. Huang. Adaptation of tandem hmms for non-speech audio event detection. *J. Acoust. Soc. Am.*, 125:2730, 2009.

[44] Su-Youn Yoon, Lisa Pierce, Amanda Huensch, Eric Juul, Samantha Perkins, Richard Sproat, and Mark Hasegawa-Johnson. Construction of a rated speech corpus of l2 learners' speech. *CALICO journal*, 2009.

[45] Lae-Hoon Kim, Mark Hasegawa-Johnson, Jun-Seok Lim, and Koeng-Mo Sung. Acoustic model for robustness analysis of optimal multipoint room equalization. *J. Acoust. Soc. Am.*, 123(4):2043–2053, 2008.

[46] Hao Tang, Yun Fu, Jilin Tu, Mark Hasegawa-Johnson, and Thomas S. Huang. Humanoid audio-visual avatar with emotive text-to-speech synthesis. *IEEE Trans. Multimedia*, 10(6):969–981, 2008.

[47] Jeung-Yoon Choi, Mark Hasegawa-Johnson, and Jennifer Cole. Finding intonational boundaries using acoustic cues related to the voice source. *Journal of the Acoustical Society of America*, 118(4):2579–88, 2005.

[48] Sarah Borys, Mark Hasegawa-Johnson, and Jennifer Cole. Prosody as a conditioning variable in speech recognition. *Illinois Journal of Undergraduate Research*, 2003.

[49] Bowon Lee, Mark Hasegawa-Johnson, and Camille Goudeseune. Open loop multichannel inversion of room impulse response. *J. Acoust. Soc. Am.*, 113(4):2202–2203, 2003.

[50] James Beauchamp, Heinrich Taube, Sever Tipei, Scott Wyatt, Lippold Haken, and Mark Hasegawa-Johnson. Acoustics, audio, and music technology education at the university of illinois. *J. Acoust. Soc. Am.*, 110(5):2961, 2001.

[51] Sumiko Takayanagi, Mark Hasegawa-Johnson, Laurie S. Eisner, and Amy Schaefer-Martinez. Information theory and variance estimation techniques in the analysis of category rating data and paired comparisons. *J. Acoust. Soc. Am.*, 102:3091, 1997.

[52] Mark A. Hasegawa-Johnson. Burst spectral measures and formant frequencies can be used to accurately discriminate stop place of articulation. *J. Acoust. Soc. Am.*, 98:2890, 1995.

[53] Mark A. Johnson. A mapping between trainable generalized properties and the acoustic correlates of distinctive features. *MIT Speech Communication Group Working Papers*, 9:94–105, 1994.

[54] Mark A. Johnson. A mapping between trainable generalized properties and the acoustic correlates of distinctive features. *J. Acoust. Soc. Am.*, 94:1865, 1993.

[55] Mark Johnson and Tomohiko Taniguchi. On-line and off-line computational reduction techniques using backward filtering in celp speech coders. *IEEE Transactions Acoustics, Speech, and Signal Processing*, 40:2090–2093, 1992.

[56] Mark A. Johnson. Analysis of durational rhythms in two poems by robert frost. *MIT Speech Communication Group Working Papers*, 8:29–42, 1992.

[57] Mark A. Johnson. Using beam elements to model the vocal fold length in breathy voicing. *J. Acoust. Soc. Am.*, 91:2420–2421, 1992.

Book Chapters

[1] Andrew Rosenberg and Mark Hasegawa-Johnson. Automatic prosody labeling and assessment. In Carlos Gussenhoven and Aoju Chen, editors, *Oxford Handbook of Language Prosody*, pages 646–656. Oxford University Press, 2021.

[2] Heejin Kim and Mark Hasegawa-Johnson. Communication improves when human or computer listeners adapt to dysarthria. In Amy Neustein and Hemant Patil, editors, *Acoustic Analysis of Pathologies: From Infancy to Young Adulthood*, pages 181–197. Walter De Gruyter, Inc., Boston/Berlin, 2020.

[3] Ken Chen, Mark Hasegawa-Johnson, and Jennifer Cole. A factored language model for prosody-dependent speech recognition. In Michael Grimm and Kristian Kroschel, editors, *Robust Speech Recognition and Understanding*, pages 319–332. INTECH Publishing, 2007.

[4] Mark Hasegawa-Johnson. Dealing with acoustic noise. part i: Spectral estimation. In *WS06*. Center for Language and Speech Processing, July 2006.

[5] Mark Hasegawa-Johnson. Dealing with acoustic noise. part ii: Beamforming. In *WS06*. Center for Language and Speech Processing, July 2006.

Conference Papers

[1] Lama Abufares, Mark Hasegawa-Johnson, and Imad L. Al-Qadi. Optimization of hot-mix asphalt rolling pattern using gpr and markov decision processes. In *TRB Annual Meeting*, 2025.

[2] SooHwan Eom, Mark Hasegawa-Johnson, and Chang D. Yoo. SiamCTC: Learning speech representations through monotonic temporal alignment. In *Proc. Interspeech*, pages 3573–3577, 8 2025.

[3] Xulin Fan, Jialu Li, Mark Hasegawa-Johnson, and Nancy McElwain. Band-split self-supervised mamba for infant-centered audio analysis. In *Proc. Interspeech*, pages 2795–2799, 8 2025.

[4] Heting Gao, Xulin Fan, and Mark Hasegawa-Johnson. SyncDiff: Diffusion-based talking head synthesis with bottlenecked temporal visual prior for improved synchronization. In *WACV Winter Conference on Applications of Computer Vision*, 3 2025.

[5] Heather Hodges, Lorraine Olson Ramig, and Mark Hasegawa-Johnson. Improving automatic speech recognition for speakers with parkinson's disease. In *Movement Disorders*, Honolulu, 10 2025.

[6] Yannan Hu, Mark Hasegawa-Johnson, and Nancy L. McElwain. Infant-directed speech prosody and real-time infant gaze at the mother's face: A machine-learning approach. In *Society for Research in Child Development (SRCD) 2025 Biennial Meeting*, Minneapolis, Minnesota, 5 2025.

[7] Maliha Jahan, Priyam Mazumdar, Thomas Thebaud, Mark Hasegawa-Johnson, Jesus Villalba, Najim Dehak, and Laureano Moro-Velazquez. Unveiling performance bias in asr systems: A study on gender, age, accent, and more. In *Proc. ICASSP*, pages 1–5, 2025.

[8] Maliha Jahan, Yinglun Sun, Priyam Mazumdar, Zsuzsanna Fagyal, Thomas Thebaud, Jesus Villalba, Mark Hasegawa-Johnson, Najim Dehak, and Laureano Moro-Velazquez. FAI: A benchmark dataset for fairness in speech technology. In *Proc. Interspeech*, pages 1343–1347, 8 2025.

[9] Jonghwan Na, Mark Hasegawa-Johnson, and Bowon Lee. Cohort-sensitive labeling: An effective approach for enhancing asr performance. In *Proc. ICASSP*, pages 1–5, 2025.

[10] Jonghwan Na, Xiuwen Zheng, Mark Hasegawa-Johnson, and Bowon Lee. Improved recognition of the speech of people with parkinson's who stutter. In *Proc. ICASSP*, pages 1–5, 2025.

[11] Bornali Phukon, Xiuwen Zheng, and Mark Hasegawa-Johnson. Aligning ASR evaluation with human and LLM judgments: Intelligibility metrics using phonetic, semantic, and NLI approaches. In *Proc. Interspeech*, pages 5708–5712, 8 2025.

[12] Kaizhi Qian, Xulin Fan, Junrui Ni, Slava Shechtman, Mark A. Hasegawa-Johnson, Chuang Gan, and Yang Zhang. ProsodyLM: Uncovering the emerging prosody processing capabilities in speech language models. In *The Conference on Language Modeling COLM*, 10 2025.

[13] Lorraine Ramig, Mark Hasegawa-Johnson, Christopher Zwilling, Heather Hodges, Clarion Mendes, and Heejin Kim. Improving automatic speech recognition (asr) for people with parkinson's disease. In *International Congress of Parkinson's Disease and Movement Disorders*, Honolulu, Hawaii, USA, 10 2025.

[14] Satwinder Singh, Qianli Wang, Zihan Zhong, Clarion Mendes, Mark Hasegawa-Johnson, Waleed Abdulla, and Seyed Reza Shahamiri. Robust cross-etiology and speaker-independent dysarthric speech recognition. In *Proc. ICASSP*, pages 1–5, 2025.

[15] Qianli Wang, Zihan Zong, Satwinder Singh, Clarion Mendes, Mark Hasegawa-Johnson, Waleed Abdulla, and Seyed Reza Shahamiri. Dysarthric speech conformer: Adaptation for sequence-to-sequence dysarthric speech recognition. In *Proc. ICASSP*, pages 1–5, 2025.

[16] Eunseop Yoon, Heesuk Yoon, Mark Hasegawa-Johnson, and Changdong Yoo. Can video llms refuse to answer? alignment for answerability in video large language models. In *Proc. International Conference on Learning Representations (ICLR)*, 2025.

[17] Hee Suk Yoon, Eunseop Yoon, Mark A. Hasegawa-Johnson, Sungwoong Kim, and Chang D. Yoo. ConfPO: Exploiting policy model confidence for critical token selection in preference optimization. In *Proc. International Conference on Machine Learning (ICML)*, 7 2025.

[18] Xiuwen Zheng, Bornali Phukon, Jonghwan Na, Ed Cutrell, Kyu Han, Mark Hasegawa-Johnson, Pan-Pan Jiang, Aadhrik Kuila, Colin Lea, Bob MacDonald, Gautam Mantena, Venkatesh Ravichandran, Leda Sari, Katrin Tomanek, Chang D. Yoo, and Chris Zwilling. The interspeech 2025 speech accessibility project challenge. In *Proc. Interspeech*, pages 3269–3273, 8 2025.

[19] Chris Zwilling, Mark Hasegawa-Johnson, Heather Hodges, Lori Ramig, Adina Bradshaw, Clarion Mendes, Heejin Kim, Alex Barkheimer, Laura Mattie, Meg Dickinson, Shawnise Carter, and Marie Channell. The speech accessibility project: Best practices for collection and curation of disordered speech. In *Proc. Interspeech*, pages 3938–3942, 8 2025.

[20] Yannan Hu, Nancy L. McElwain, and Mark Hasegawa-Johnson. Real-time relations between fundamental frequency of infant-directed speech and infant attention at 3 months. In *24th International Congress of Infant Studies (ICIS)*, Glasgow, Scotland, July 2024. Poster presentation.

[21] Heting Gao, Mark Hasegawa-Johnson, and Chang D. Yoo. G2pu: Grapheme-to-phoneme transducer with speech units. In *Proc. ICASSP*, 2024.

[22] Heting Gao, Kaizhi Qian, Junrui Ni, Chuang Gan, Mark A. Hasegawa-Johnson, Shiyu Chang, and Yang Zhang. Speech self-supervised learning using diffusion model synthetic data. In *Proc. International Conference on Machine Learning (ICML)*, 7 2024.

[23] Yannan Hu, Nancy L. McElwain, and Mark Hasegawa-Johnson. Real-time relations between prosodic features of infant-directed speech and infant attention at 3 months. In *Speech Prosody*, pages 220–224, 7 2024.

[24] Maliha Jahan, Helin Wang, Thomas Thebaud, Yinglun Sun, Giang Le, Zsuzsanna Fagyal, Odette Scharenborg, Mark Hasegawa-Johnson, Laureano Moro Velazquez, and Najim Dehak. Finding spoken identifications: Using gpt-4 annotation for an efficient and fast dataset creation pipeline. In *LREC-COLING*, pages 7296–7306, 5 2024.

[25] Mohammad Nur Hossain Khan, Nancy McElwain, Mark Hasegawa-Johnson, and Bashima Islam. Infantmotion2vec: Unlabeled data-driven infant pose estimation using a single chest IMU. In *IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN 2024)*, 10 2024.

[26] Heejin Kim, Clarion Mendes, Mark Hasegawa-Johnson, Meg Dickinson, and Erik Hege. An overview of speech data in the speech accessibility project: Speech variations for inclusive technology. In *American Speech Language Hearing Association (ASHA) Annual Convention*, 12 2024. unpublished presentation.

[27] Jialu Li, Mark Hasegawa-Johnson, and Karrie Karahalios. Enhancing child vocalization classification with phonetically-tuned embeddings for assisting autism diagnosis. In *Proc. Interspeech*, 9 2024.

[28] Satwinder Singh, Zihan Zhong, Qianli Wang, Clarion Mendes, Mark Hasegawa-Johnson, Waleed Abdulla, and Seyed Reza Shahamiri. A comprehensive performance evaluation of whisper models in dysarthric speech recognition. In *International Conference on Neural Information Processing (ICONIP2024)*, 12 2024.

[29] Liming Wang, Mark Hasegawa-Johnson, and Chang Yoo. Unsupervised speech recognition with n-skipgram and positional unigram matching. In *Proc. ICASSP*, 2024.

[30] Junkai Wu, Xulin Fan, Bo-Ru Lu, Xilin Jiang, Nima Mesgarani, Mark Hasegawa-Johnson, and Mari Ostendorf. Just asr + llm? a study on speech large language models' ability to identify and understand speaker in spoken dialogue. In *IEEE Spoken Language Technology (SLT)*, 12 2024.

[31] Eunseop Yoon, Hee Suk Yoon, John Harvill, Mark Hasegawa-Johnson, and Chang D. Yoo. Li-tta: Language informed test-time adaptation for automatic speech recognition. In *Proc. Interspeech*, 9 2024.

[32] Hee Suk Yoon, Eunseop Yoon, Joshua Tian Jin Tee, Mark A. Hasegawa-Johnson, Yingzhen Li, and Chang D. Yoo. C-TPT: Calibrated test-time prompt tuning for vision-language models via text feature dispersion. In *Proc. International Conference on Learning Representations (ICLR)*, 5 2024.

[33] Xiuwen Zheng, Bornali Phukon, and Mark Hasegawa-Johnson. Fine-tuning automatic speech recognition for people with parkinson's: An effective strategy for enhancing speech technology accessibility. In *Proc. Interspeech*, 9 2024.

[34] Kai Chieh Chang, Mark Hasegawa-Johnson, Nancy L McElwain, and Bashima Islam. Classification of infant sleep/wake states: cross-attention among large scale pretrained transformer networks using audio, ecg, and imu data. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 2370–2377. IEEE, 2023.

[35] Wonjune Kang, Mark Hasegawa-Johnson, and Deb Roy. End-to-end zero-shot voice conversion with location-variable convolutions. In *Proc. Interspeech*, 2023.

[36] Jialu Li, Mark Hasegawa-Johnson, and Nancy McElwain. Towards robust family-infant audio analysis based on unsupervised pretraining of wav2vec 2.0 on large-scale unlabeled family audio. In *Proc. Interspeech*, 2023.

[37] Liming Wang, Mark Hasegawa-Johnson, and Chang D. Yoo. A theory of unsupervised speech recognition. In *ACL*, 7 2023.

[38] Liming Wang, Junrui Ni, Heting Gao, Jialu Li, Kai Chieh Chang, Xulin Fan, Junkai Wu, Mark Hasegawa-Johnson, and Chang D. Yoo. Speak and decipher and sign: Toward unsupervised speech-to-sign language recognition. In *Findings of ACL*, 7 2023.

[39] Zhongweiyang Xu, Xulin Fan, and Mark Hasegawa-Johnson. Dual-path cross-modal attention for better audio-visual speech extraction. In *Proceedings of ICASSP*, 2023. Recognized as one of the top 3% of papers at the conference.

[40] Eunseop Yoon, Hee Suk Yoon, Dhananjaya Gowda, SooHwan Eom, Daehyeok Kim, John Harvill, Heting Gao, Mark Hasegawa-Johnson, Chanwoo Kim, and Chang D. Yoo. Mitigating the exposure bias in sentence-level grapheme-to-phoneme (G2P) transduction. In *Proc. Interspeech*, 2023.

[41] Wanyue Zhai and Mark Hasegawa-Johnson. Wav2ToBI: a new approach to automatic ToBI transcription. In *Proc. Interspeech*, 2023.

[42] Chak Ho Chan, Kaizhi Qian, Yang Zhang, and Mark Hasegawa-Johnson. Speechsplit2.0: Unsupervised speech disentanglement for voice conversion without tuning autoencoder bottlenecks. In *ICASSP*, pages 6332–6336, 2022.

[43] Heting Gao, Junrui Ni, Kaizhi Qian, Yang Zhang, Shiyu Chang, and Mark Hasegawa-Johnson. WavPrompt: Towards Few-Shot Spoken Language Understanding with Frozen Language Models. In *Proc. Interspeech 2022*, pages 2738–2742, 2022.

[44] John Harvill, Roxana Girju, and Mark Hasegawa-Johnson. Syn2Vec: Synset colexification graphs for lexical semantic similarity. In *Proc. NAACL*, page 5259–5270, 2022.

[45] John Harvill, Mark Hasegawa-Johnson, and Chang D. Yoo. Frame-Level Stutter Detection. In *Proc. Interspeech 2022*, pages 2843–2847, 2022.

[46] John Harvill, Yash Wani, Narendra Ahuja, Mark Hasegawa-Johnson, David Chestek, Mustafa Alam, and David Beiser. Estimation of respiratory rate from breathing audio. In *44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2022.

[47] Haeyong Kang, Rusty John Lloyd Mina, Sultan Rizky Hikmawan Madjid, Jaehong Yoon, Mark Hasegawa-Johnson, Sung Ju Hwang, and Chang D Yoo. Forget-free continual learning with winning subnetworks. In *Proc. International Conference on Machine Learning (ICML)*, volume 162, pages 10734–10750, 6 2022.

[48] Mahir Morshed and Mark Hasegawa-Johnson. Cross-lingual articulatory feature information transfer for speech recognition using recurrent progressive neural networks. In *Proc. Interspeech 2022*, pages 2298–2302, 2022.

[49] Junrui Ni, Liming Wang, Heting Gao, Kaizhi Qian, Yang Zhang, Shiyu Chang, and Mark Hasegawa-Johnson. Unsupervised Text-to-Speech Synthesis by Unsupervised Automatic Speech Recognition. In *Proc. Interspeech 2022*, pages 461–465, 2022.

[50] Kaizhi Qian, Yang Zhang, Heting Gao, Junrui Ni, C.-I. Lai, David Cox, Mark Hasegawa-Johnson, and Shiyu Chang. Contentvec: An improved selfsupervised speech representation by disentangling speakers. In *Proc. International Conference on Machine Learning (ICML)*, pages 18003–18017, 2022.

[51] Leda Sarı, Mark Hasegawa-Johnson, and Samuel Thomas. Auxiliary networks for joint speaker adaptation and speaker change detection. In *IEEE Transactions on Audio, Speech, and Language*, volume 29, pages 324–333, 2022.

[52] Liming Wang, Siyuan Feng, Mark A. Hasegawa-Johnson, and Chang D. Yoo. Self-supervised semantic-driven phoneme discovery for zero-resource speech recognition. In *ACL*, page 8027–8047, 5 2022.

[53] Raymond Yeh, Mark Hasegawa-Johnson, and Alexander Schwing. Equivariance discovery by learned parameter-sharing. In *AISTATS*, 2022.

[54] Hee Suk Yoon, Eunseop Yoon, John Harvill, Sunjae Yoon, Mark Hasegawa-Johnson, and Chang D. Yoo. SMSMix: Sense maintained sentence mixup for word sense disambiguation. In *EMNLP*, page 1493–1502, 12 2022.

[55] Siyuan Feng, Piotr Żelasko, Laureano Moro-Velázquez, Ali Abavisani, Mark Hasegawa-Johnson, Odette Scharenborg, and Najim Dehak. How phonotactics affect multilingual and zero-shot asr performance. In *Proc. ICASSP*, pages 7238–7242, 2021.

[56] Heting Gao, Junrui Ni, Yang Zhang, Kaizhi Qian, Shiyu Chang, and Mark Hasegawa-Johnson. Zero-shot cross-lingual phonetic recognition with external language embedding. In *Proc. Interspeech*, pages 1304–1308, 2021.

[57] John Harvill, Dias Issa, Mark Hasegawa-Johnson, and Changdong Yoo. Synthesis of new words for improved dysarthric speech recognition on an expanded vocabulary. In *Proc. ICASSP*, pages 6428–6432, 2021.

[58] John Harvill, Yash R. Wani, Mark Hasegawa-Johnson, Narendra Ahuja, David Beiser, and David Chestek. Classification of COVID-19 from Cough Using Autoregressive Predictive Coding Pretraining and Spectral Data Augmentation. In *Proc. Interspeech*, pages 926–930, 2021.

[59] Kaizhi Qian, Yang Zhang, Shiyu Chang, Chuang Gan, David D. Cox, Mark Hasegawa-Johnson, and Jinjun Xiong. Global rhythm style transfer without text transcriptions. In *ICML*, 2021.

[60] Kiran Ramnath, Leda Sarı, Mark Hasegawa-Johnson, and Chang Yoo. Worldly wise (wow) - cross-lingual knowledge fusion for fact-based visual spoken-question answering. In *Proc. NAACL*, page 1908–1919, 2021.

[61] Hui Shi, Yang Zhang, Hao Wu, Shiyu Chang, Kaizhi Qian, Mark Hasegawa-Johnson, and Jishen Zhao. Continuous cnn for nonuniform time series. In *Proc. ICASSP*, 2021.

[62] Liming Wang, Xinsheng Wang, Mark Hasegawa-Johnson, Odette Scharenborg, and Najim Dehak. Align or attend? toward more efficient and accurate spoken word discovery using speech-to-image retrieval. In *Proc. ICASSP*, 2021.

[63] Xinsheng Wang, Siyuan Feng, Jihua Zhu, Mark Hasegawa-Johnson, and Odette Scharenborg. Show and speak: Directly synthesize spoken description of images. In *Proc. ICASSP*, 2021.

[64] Zhonghao Wang, Mo Yu, Kai Wang, Jinjun Xiaong, Wen mei Hwu, Mark Hasegawa-Johnson, and Humphrey Shi. Interpretable visual reasoning via induced symbolic space. In *ICCV*, pages 1878–1887, 2021.

[65] Junzhe Zhu, Mark Hasegawa-Johnson, and Nancy McElwain. A comparison study on infant-parent voice diarization. In *Proc. ICASSP*, pages 7178–7182, 2021.

[66] Junzhe Zhu, Raymond Yeh, and Mark Hasegawa-Johnson. Multi-decoder dprnn: Source separation for variable number of speakers. In *Proc. ICASSP*, pages 3420–3424, 2021.

[67] Ali Abavisani and Mark Hasegawa-Johnson. Automatic estimation of intelligibility measure for consonants in speech. In *Proc. Interspeech*, pages 1161–1165, 2020.

[68] Mark Hasegawa-Johnson. Multimodal distant supervision. In *NeurIPS Workshop on Self-Supervised Learning for Speech and Audio*, 2020.

[69] Jialu Li and Mark Hasegawa-Johnson. Autosegmental neural nets: Should phones and tones be synchronous or asynchronous? In *Proc. Interspeech*, pages 1027–1031, 2020.

[70] Daniel Morrow, Renato F.L. Azevedo, Leda Sari, Kuangxiao Gu, Tarek Sakakini, Mark Hasegawa-Johnson, Suma Bhat, James Graumlich, Thomas Huang, Andrew Hariharan, Yunxin Shao, and Elizabeth Cox. Closing the loop in computer agent/patient communication. In *Proceedings of the 2020 Human Factors and Ergonomics Society Annual Meeting*, Chicago, IL, 2020.

[71] Kaizhi Qian, Zeyu Jin, Mark Hasegawa-Johnson, and Gautham Mysore. F0-consistent many-to-many non-parallel voice conversion via conditional autoencoder. In *Proc. ICASSP*, pages 6284–6288, 2020.

[72] Kaizhi Qian, Yang Zhang, Shiyu Chang, Mark Hasegawa-Johnson, and David Cox. Unsupervised speech decomposition via triple information bottleneck. In *Proc. International Conference on Machine Learning (ICML)*, volume 119, pages 7836–7846, 2020.

[73] Tarek Sakakini, Jong Yoon Lee, Aditya Srinivasa, Renato Azevedo, Victor Sadauskas, Kuangxiao Gu, Suma Bhat, Dan Morrow, James Graumlich, Saqib Walayat, Mark Hasegawa-Johnson, Donald Wilpern, and Ann Willemsen-Dunlap. Automatic text simplification of health materials in low-resource domains. In *LOUHI: 11th International Workshop on Health Text Mining and Information Analysis*, 2020.

[74] Leda Sarıand Mark Hasegawa-Johnson. Deep F-Measure Maximization for End-to-End Speech Understanding. In *Proc. Interspeech*, pages 1580–1584, 2020.

[75] Leda Sarı, Samuel Thomas, and Mark Hasegawa-Johnson. Training spoken language understanding systems with non-parallel speech and text. In *Proc. ICASSP*, pages 8109–8113, 2020.

[76] Justin van der Hout, Mark Hasegawa-Johnson, and Odette Scharenborg. Evaluating automatically generated phoneme captions for images. In *Proc. Interspeech*, pages 2317–2321, 2020.

[77] Liming Wang and Mark Hasegawa-Johnson. A DNN-HMM-DNN Hybrid Model for Discovering Word-Like Units from Spoken Captions and Image Regions. In *Proc. Interspeech*, pages 1456–1460, 2020.

[78] Junzhe Zhu, Mark Hasegawa-Johnson, and Leda Sari. Identify speakers in cocktail parties with end-to-end attention. In *Proc. Interspeech*, pages 3092–3096, 2020.

[79] Piotr Żelasko, Laureano Moro-Velázquez, Mark Hasegawa-Johnson, Odette Scharenborg, and Najim Dehak. That Sounds Familiar: An Analysis of Phonetic Representations Transfer Across Languages. In *Proc. Interspeech 2020*, pages 3705–3709, 2020.

[80] Renato F.L. Azevedo, Dan Morrow, Kuangxiao Gu, Thomas Huang, Mark Hasegawa-Johnson, P. Soni, S. Tang, Tarek Sakakini, Suma Bhat, Ann Willemsen-Dunlap, and James Graumlich. The influence of computer agent characteristics on user preferences in health contexts. In *Proceedings of the 2019 Human Factors and Ergonomics Society Health Care Symposium*, 2019.

[81] Mark Hasegawa-Johnson, Najim Dehak, and Odette Scharenborg. Position paper: Indirect supervision for dialog systems in unwritten languages. In *International Workshop on Spoken Dialog Systems*, 2019.

[82] Di He, Xuesong Yang, Boon Pang Lim, Yi Liang, Mark Hasegawa-Johnson, and Deming Chen. When ctc training meets acoustic landmarks. In *ICASSP*, pages 5996–6000, 2019.

[83] Laureano Moro-Velazquez, JaeJin Cho, Shinji Watanabe, Mark A. Hasegawa-Johnson, Odette Scharenborg, Heejin Kim, and Najim Dehak. Study of the Performance of Automatic Speech Recognition Systems in Speakers with Parkinson's Disease. In *Proc. Interspeech 2019*, pages 3875–3879, 2019.

[84] Mary Pietrowicz, Carla Agurto, Jonah Casebeer, Mark Hasegawa-Johnson, Karrie Karahalios, and Guillermo Cecchi. Dimensional analysis of laughter in female conversational speech. In *Proc. ICASSP*, pages 6600–6604, 2019.

[85] Leda Sari, Samuel Thomas, Mark Hasegawa-Johnson, and Michael Picheny. Pre-training of speaker embeddings for low-latency speaker change detection in broadcast news. In *Proc. ICASSP*, pages 3093:1–5, 2019.

[86] Leda Sarı, Samuel Thomas, and Mark A. Hasegawa-Johnson. Learning Speaker Aware Offsets for Speaker Adaptation of Neural Networks. In *Proc. Interspeech 2019*, pages 769–773, 2019.

[87] Odette Scharenborg, Jiska Koemans, Cybelle Smith, Mark A. Hasegawa-Johnson, and Kara D. Federmeier. The neural correlates underlying lexically-guided perceptual learning. In *Proc. Interspeech*, pages 1223–1227, 2019.

[88] Liming Wang and Mark A. Hasegawa-Johnson. Multimodal word discovery and retrieval with phone sequence and image concepts. In *Proc. Interspeech*, pages 2683–2687, 2019.

[89] Renato Azevedo, Daniel G. Morrow, Kuangxiao Gu, Thomas Huang, Mark Allan Hasegawa-Johnson, James Graumlich, Victor Sadauskas, Tarek J. Sakakini, Suma Pallathadka Bhat, Ann M. Willemsen-Dunlap, and Donald J. Halpin. Computer agents and patient memory for medication information. In *APA Annual Meeting*, 2018.

[90] Renato F. L. Azevedo, Dan Morrow, James Graumlich Ann Willemsen-Dunlap Mark Hasegawa-Johnson Thomas S. Huang, Kuangxiao Gu, Suma Bhat, Tarek Sakakini, Victor Sadauskas, and Donald J. Halpin. Using conversational agents to explain medication instructions to older adults. In *AMIA Annu Symp Proc.*, page 185–194, 2018.

[91] Wenda Chen, Mark Hasegawa-Johnson, and Nancy Chen. Recognizing zero-resourced languages based on mismatched machine transcriptions. In *Proc. ICASSP*, pages 5979–5983, 2018.

[92] Wenda Chen, Mark Hasegawa-Jonson, and Nancy F.Y. Chen. Topic and keyword identification for low-resourced speech using cross-language transfer learning. In *Proc. Interspeech*, pages 2047–2051, 2018.

[93] Amit Das and Mark Hasegawa-Johnson. Improving dnns trained with non-native transcriptions using knowledge distillation and target interpolation. In *Proc. Interspeech*, pages 2434–2438, 2018.

[94] Mark Hasegawa-Johnson, Alan Black, Lucas Ondel, Odette Scharenborg, and Francesco Ciannella. Image2speech: Automatically generating audio descriptions of images. In *Journal of the International Science and General Applications (ISGA)*, volume 1, 2018.

[95] Di He, Boon Pang Lim, Xuesong Yang, Mark Hasegawa-Johnson, and Deming Chen. Improved asr for under-resourced languages through multi-task learning with acoustic landmarks. In *Proc. Interspeech*, pages 2618–2622, 2018.

[96] Jialu Li and Mark Hasegawa-Johnson. A comparable phone set for the timit dataset discovered in clustering of listen, attend and spell. In *NeurIPS Workshop on Interpretability and Robustness in Audio, Speech, and Language*, 2018.

[97] Teck Yian Lim, Raymond Yeh, Yijia Xu, Minh Do, and Mark Hasegawa-Johnson. Time-frequency networks for audio super-resolution. In *Proc. ICASSP*, 2018.

[98] Lucas Ondel, Pierre Godard, Laurent Besacier, Elin Larsen, Mark Hasegawa-Johnson, Odette Scharenborg, Emmanuel Dupoux, Lukas Burget, François Yvon, and Sanjeev Khudanpur. Bayesian models for unit discovery on a very low resource language. In *Proc. ICASSP*, pages 5939–5943, 2018.

[99] Kaizhi Qian, Yang Zhang, Shiyu Chang, Xuesong Yang, Dinei Florencio, and Mark Hasegawa-Johnson. Deep learning based speech beamforming. In *Proc. ICASSP*, pages 5389–5393, 2018.

[100] Leda Sari and Mark Hasegawa-Johnson. Speaker adaptation with an auxiliary network. In *MLSLP (ISCA Workshop on Machine Learning for Speech and Language Processing)*, 2018.

[101] Leda Sari, Mark Hasegawa-Johnson, S. Kumaran, Georg Stemmer, and N. Nair Krishnakumar. Speaker adaptive audio-visual fusion for the open-vocabulary section of avicar. In *Proc. Interspeech*, pages 3524–3528, 2018.

[102] Odette Scharenborg, Laurent Besacier, Alan Black, Mark Hasegawa-Johnson, Florian Metze, Graham Neubig, Sebastian Stüker, Pierre Godard, Markus Müller, Lucas Ondel, Shruti Palaskar, Philip Arthur, Francesco Ciannella, Mingxing Du, Elin Larsen, Danny Merkx, Rachid Riad, Liming Wang, and Emmanuel Dupoux. Linguistic unit discovery from multi-modal inputs in unwritten languages: Summary of the speaking rosetta jsalt 2017 workshop. In *Proc. ICASSP*, 2018.

[103] Odette Scharenborg, Patrick Ebel, Francesco Ciannella, Mark Hasegawa-Johnson, and Najim Dehak. Building an asr system for mboshi using a cross-language definition of acoustic units approach. In *Proc. SLTU (Speech and Language Technology for Under-resourced languages)*, pages 167–171, 2018.

[104] Odette Scharenborg, Sebastian Tiesmeyer, Mark Hasegawa-Johnson, and Najim Dehak. Visualizing phoneme category adaptation in deep neural networks. In *Proc. Interspeech*, pages 1482–1486, 2018.

[105] Yijia Xu, Mark Hasegawa-Johnson, and Nancy L. McElwain. Infant emotional outbursts detection in infant-parent spoken interactions. In *Proc. Interspeech*, pages 242–246, 2018.

[106] Xuesong Yang, Kartik Audhkhasi, Andrew Rosenberg, Samuel Thomas, Bhuvana Ramabhadran, and Mark Hasegawa-Johnson. Joint modeling of accents and acoustics for multi-accent speech recognition. In *Proc. ICASSP*, pages 5989–5993, 2018.

[107] Raymond A. Yeh, Teck Yian Lim, Chen Chen, Alexander G. Schwing, Mark Hasegawa-Johnson, and Minh N. Do. Image restoration with deep generative models. In *Proc. IEEE ICASSP*, pages 6772–6772, 2018.

[108] Di He, Boon Pang Lim, Xuesong Yang, Mark Hasegawa-Johnson, and Deming Chen. Selecting frames for automatic speech recognition based on acoustic landmarks. In *Acoustics 17*, Boston, June 2017.

[109] Xiang Kong, Xuesong Yang, Jeung-Yoon Choi, Mark Hasegawa-Johnson, and Stefanie Shattuck-Hufnagel. Landmark-based consonant voicing detection on multilingual corpora. In *Acoustics 17*, Boston, June 2017.

[110] Xiang Kong, Xuesong Yang, Jeung-Yoon Choi, Mark Hasegawa-Johnson, and Stefanie Shattuck-Hufnagel. Landmark-based consonant voicing detection on multilingual corpora. In *Acoustics 17*, Boston, June 2017.

[111] Shiyu Chang, Yang Zhang, Wei Han, Mo Yu andXiaoxiao Guo, Wei Tan, Xiaodong Cui, Michael Witbrock, Mark Hasegawa-Johnson, and Thomas Huang. Dilated recurrent neural networks. In *NIPS*, 2017.

[112] Shiyu Chang, Yang Zhang, Jiling Tang, Dawei Yin, Yi Chang, Mark Hasegawa-Johnson, and Thomas Huang. Streaming recommender systems. In *WWW 2017*, pages 381–389, 2017.

[113] Wenda Chen, Mark Hasegawa-Johnson, Nancy F. Chen, and Boon Pang Lim. Mismatched crowdsourcing from multiple annotator languages for recognizing zero-resourced languages: A nullspace clustering approach. In *Proc. Interspeech*, pages 2789–2793, 2017.

[114] Amit Das, Mark Hasegawa-Johnson, and Karel Vesely. Deep autoencoder based multi-task learning using probabilistic transcription. In *Proc. Interspeech*, pages 2073–2077, 2017.

[115] Van Hai Do, Nancy F. Chen, Boon Pang Lim, and Mark Hasegawa-Johnson. Multi-Task Learning Using Mismatched Transcription for Under-Resourced Speech Recognition. In *Proc. Interspeech 2017*, pages 734–738, 2017.

[116] Mark Hasegawa-Johnson, Alan Black, Lucas Ondel, Odette Scharenborg, and Francesco Ciannella. Image2speech: Automatically generating audio descriptions of images. In *Proc. Internat. Conference on Natural Language, Signal and Speech Processing (ICNLSSP)*, Casablanca, Morocco, 2017.

[117] Di He, Zuofu Cheng, Mark Hasegawa-Johnson, and Deming Chen. Using approximated auditory roughness as a pre-filtering feature for human screaming and affective speech aed. In *Proc. Interspeech*, pages 1914–1918, 2017.

[118] Preethi Jyothi and Mark Hasegawa-Johnson. Low-resource grapheme-to-phoneme conversion using recurrent neural networks. In *Proc. ICASSP*, pages 5030–5034, 2017.

[119] Pavlos Papadopoulos, Ruchir Travadi, Colin Vaz, Nikolaos Malandrakis, Ulf Hermjakob, Nima Pourdamghani, Michael Pust, Boliang Zhang, Xiaoman Pan, Di Lu, Ying Lin, Ondrej Glembek, Murali Karthick B, Martin Karafiat, Lukas Burget, Mark Hasegawa-Johnson, Heng Ji, Jonathan May, Kevin Knight, and Shrikanth Narayanan. Team ELISA system for DARPA LORELEI speech evaluation 2016. In *Proc. Interspeech*, pages 2053–2057, 2017.

[120] Mary Pietrowicz, Mark Hasegawa-Johnson, and Karrie Karahaliqos. Discovering dimensions of perqceived vocal expression in semi-structured, unscripted oral history accoqunts. In *Proc. ICASSP*, pages 2901:1–4, 2017.

[121] Kaizhi Qian, Yang Zhang, Shiyu Chang, Xuesong Yang, Dinei Florencio, and Mark Hasegawa-Johnson. Speech enhancement using bayesian wavenet. In *Proc. Interspeech*, pages 2013–2017, 2017.

[122] Odette Scharenborg, Francesco Ciannella, Shruti Palaskar, Alan Black, Florian Metze, Lucas Ondel, and Mark Hasegawa-Johnson. Building an asr system for a low-resource language through the adaptation of a high-resource language asr system: Preliminary results. In *Proc. Internat. Conference on Natural Language, Signal and Speech Processing (ICNLSSP)*, Casablanca, Morocco, 2017.

[123] Raymond Yeh, Chen Chen, Teck Yian Lim, Alexander G. Schwing, Mark Hasegawa-Johnson, and Minh N. Do. Semantic image inpainting with deep generative networks. In *CVPR*, pages 5485–5493, 2017.

[124] Yang Zhang, Dinei Florêncio, and Mark Hasegawa-Johnson. Glottal Model Based Speech Beamforming for ad-hoc Microphone Arrays. In *Proc. Interspeech 2017*, pages 2675–2679, 2017.

[125] Shiyu Chang, Yang Zhang, Jiliang Tang, Dawei Lin, Yi Chang, Mark Hasegawa-Johnson, and Thomas Huang. Positive-unlabeled learning in streaming networks. In *KDD*, pages 755–764, 2016.

[126] Wenda Chen, Mark Hasegawa-Johnson, Nancy Chen, Preethi Jyothi, and Lav Varshney. Clustering-based phonetic projection in mismatched crowdsourcing channels for low-resourced asr. In *WSSAP (Workshop on South and Southeast Asian Natural Language Processing)*, pages 133–141, 2016.

[127] Amit Das and Mark Hasegawa-Johnson. An investigation on training deep neural networks using probabilistic transcription. In *Proc. Interspeech*, pages 3858–3862, 2016.

[128] Amit Das, Preethi Jyothi, and Mark Hasegawa-Johnson. Automatic speech recognition using probabilistic transcriptions in swahili, amharic and dinka. In *Proc. Interspeech*, pages 3524–3528, 2016.

[129] Van Hai Do, Nancy F. Chen, Boon Pang Lim, and Mark Hasegawa-Johnson. Analysis of mismatched transcriptions generated by humans and machines for under-resourced languages. In *Proc. Interspeech*, pages 3863–3867, 2016.

[130] Van Hai Do, Nancy F. Chen, Boon Pang Lim, and Mark Hasegawa-Johnson. A many-to-one phone mapping approach for cross-lingual speech recognition. In *12th IEEE-RIVF International Conference on Computing and Communication Technologies*, pages 120–124, Hanoi, Vietnam, 11 2016.

[131] Van Hai Do, Nancy F. Chen, Boon Pang Lim, and Mark Hasegawa-Johnson. Speech recognition of under-resourced languages using mismatched transcriptions. In *International Conference on Asian Language Processing IALP*, Tainan, Taiwan, 11 2016.

[132] Chunxi Liu, Preethi Jyothi, Hao Tang, Vimal Manohar, Rose Sloan, Tyler Kekona, Mark Hasegawa-Johnson, and Sanjeev Khudanpur. Adapting asr for under-resourced languages using mismatched transcriptions. In *Proc. ICASSP*, pages 5840–5844, 2016.

[133] Kaizhi Qian, Yang Zhang, and Mark Hasegawa-Johnson. Application of local binary patterns for svm based stop consonant detection. In *Speech Prosody*, pages 1114–1118, 2016.

[134] Ruobai Wang, Yang Zhang, Zhijian Ou, and Mark Hasegawa-Johnson. Use of particle filtering and mcmc for inference in probabilistic acoustic tube model. In *IEEE Workshop on Statistical Signal Processing*, 2016.

[135] Yanlu Xie, Mark Hasegawa-Johnson, Leyuan Qu, and Jinsong Zhang. Landmark of mandarin nasal codas and its application in pronunciation error detection. In *Proc. ICASSP*, 2016.

[136] Xuesong Yang, Xiang Kong, Mark Hasegawa-Johnson, and Yanlu Xie. Landmark-based pronunciation error identification on L2 Mandarin Chinese. In *Speech Prosody*, pages 247–251, 2016.

[137] Raymond Yeh, Mark Hasegawa-Johnson, and Minh Do. Stable and symmetric filter convolutional neural network. In *Proc. ICASSP*, pages 2652–2656, 2016.

[138] Yang Zhang, Gautham Mysore, Florian Berthouzoz, and Mark Hasegawa-Johnson. Analysis of prosody increment induced by pitch accents for automatic emphasis correction. In *Speech Prosody*, pages 79–83, 2016.

[139] Amit Das and Mark Hasegawa-Johnson. Cross-lingual transfer learning during supervised training in low resource scenarios. In *Proc. Interspeech*, pages 3531–3535, 2015.

[140] Preethi Jyothi and Mark Hasegawa-Johnson. Acquiring speech transcriptions using mismatched crowdsourcing. In *Proc. AAAI*, pages 1263–1269, 2015.

[141] Preethi Jyothi and Mark Hasegawa-Johnson. Improving hindi broadcast asr by adapting the language model and pronunciation model using a priori syntactic and morphophonemic knowledge. In *Proc. Interspeech*, pages 3164–3168, 2015.

[142] Preethi Jyothi and Mark Hasegawa-Johnson. Transcribing continuous speech using mismatched crowdsourcing. In *Proc. Interspeech 2015*, pages 2774–2778, 2015.

[143] Preethi Jyothi and Mark Hasegawa-Johnson. Transcribing continuous speech using mismatched crowdsourcing. In *Proc. Interspeech*, pages 2774–2778, 2015.

[144] Mary Pietrowicz, Mark Hasegawa-Johnson, and Karrie Karahalios. Acoustic correlates for perceived effort levels in expressive speech. In *Proc. Interspeech*, pages 3720–3724, 2015.

[145] Jia-Chen Ren, Lawrence Angrave, and Mark Hasegawa-Johnson. Classtranscribe: A new tool with new educational opportunities for student crowdsourced college lecture transcriptions. In *SLaTE (the Workshop on Speech and Language Technology in Education)*, 2015.

[146] Jia Chen Ren, Mark Hasegawa-Johnson, and Lawrence Angrave. Classtranscribe. In *ICER Conference*, 2015.

[147] Yang Zhang, Zhijian Ou, and Mark Hasegawa-Johnson. Incorporating am-fm effect in voiced speech for probabilistic acoustic tube model. In *Proc. WASPAA*, 2015.

[148] Xiayu Chen, Yang Zhang, and Mark Hasegawa-Jonson. An iterative approach to decision tree training for context dependent speech synthesis. In *Proc. Interspeech*, pages 2327–2331, 2014.

[149] Po-Sen Huang, Minje Kim, Mark Hasegawa-Johnson, and Paris Smaragdis. Singing-voice separation from monaural recordings using deep recurrent neural networks. In *Proceedings of ISMIR*, 2014.

[150] Preethi Jyothi, Jennifer Cole, Mark Hasegawa-Johnson, and Vandana Puri. An Investigation of Prosody in Hindi Narrative Speech. In *Proc. Speech Prosody 2014*, pages 623–627, 2014.

[151] Alina Khasanova, Jennifer Cole, and Mark Hasegawa-Johnson. Detecting articulatory compensation in acoustic data through linear regression modeling. In *Proc. Interspeech 2014*, pages 925–929, 2014.

[152] Kai-Hsiang Lin, Pooya Khorrami, Jiangping Wang, Mark Hasegawa-Johnson, and Thomas S. Huang. Foreground object detection in highly dynamic scenes using saliency. In *Proceedings of ICIP*, pages 1125–1129, 2014.

[153] Yang Zhang, Zhijian Ou, and Mark Hasegawa-Johnson. Improvement of probabilistic acoustic tube model for speech decomposition. In *ICASSP*, 2014.

[154] Elabbas Benmamoun and Mark Hasegawa-Johnson. How different are arabic dialects from each other and from classical arabic. In *6th Annual Arabic Linguistics Symposium*, Ifrane, Morocco, June 2013.

[155] Mahmoud Abunasser, Abbas Benmamoun, and Mark Hasegawa-Johnson. Pronunciation variation metric for four dialects of arabic. In *AIDA 10 (Association Internationale de Dialectologie Arabe)*, Qatar University, 2013.

[156] Galen Andrew, Raman Arora, Sujeeth Bharadwaj, Jeff Bilmes, Mark Hasegawa-Johnson, and Karen Livescu. Using articulatory measurements to learn better acoustic features. In *Proc. Workshop on Speech Production in Automatic Speech Recognition*, Lyon, France, 2013.

[157] Sujeeth Bharadwaj, Mark Hasegawa-Johnson, Jitendra Ajmera, Om Deshmukh, and Ashish Verma. Sparse hidden markov models for purer clusters. In *Proc. ICASSP*, 2013.

[158] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. Automatic long audio alignment for conversational arabic speech. In *Qatar Foundation Annual Research Conference*, 2013.

[159] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. Development of a spontaneous large vocabulary speech recognition system for qatari arabic. In *Qatar Foundation Annual Research Conference*, 2013.

[160] Po-Sen Huang, Li Deng, Mark Hasegawa-Johnson, and Xiaodong He. Random features for kernel deep convex network. In *Proc. ICASSP*, pages 8096–8900, 2013.

[161] Amit Juneja and Mark Hasegawa-Johnson. Experiments on context-awareness and phone error propagation in human and machine speech recognition. In *Proc. Workshop on Speech Production in Automatic Speech Recognition*, Lyon, France, 2013.

[162] Mark Hasegawa-Johnson, Elabbas Benmamoun, Eiman Mustafawi, Mohamed Elmahdy, and Rehab Duwairi. On the definition of the word 'segmental'. In *Speech Prosody*, pages 159–162, 2012.

[163] Mark Hasegawa-Johnson, Elabbas Benmamoun, Eiman Mustafawi, Mohamed Elmahdy, and Rehab Duwairi. On the definition of the word "segmental". In *Proc. Speech Prosody 2012*, pages 159–162, 2012.

[164] Po-Sen Huang, Mark Hasegawa-Johnson, Wotao Yin, and Tom Huang. Opportunistic sensing: Unattended acoustic sensor selection using crowdsourcing models. In *IEEE Workshop on Machine Learning in Signal Processing*, 2012.

[165] Po-Sen Huang, Jianchao Yang, Mark Hasegawa-Johnson, Feng Liang, and Thomas S. Huang. Pooling robust shift-invariant sparse representations of acoustic signals. In *Proc. Interspeech*, pages 2518–2521, 2012.

[166] Tim Mahrt, Jennifer Cole, Margaret Fleck, and Mark Hasegawa-Johnson. F0 and the perception of prominence. In *Proc. Interspeech 2012*, pages 2422–2425, 2012.

[167] Jui-Ting Huang, Mark Hasegawa-Johnson, and Jennifer Cole. How unlabeled data change the acoustic models for phonetic classification. In *Workshop on New Tools and Methods for Very Large Scale Phonetics Research*, University of Pennsylvania, Jan. 2011.

[168] Tim Mahrt, Jui-Ting Huang, Yoonsook Mo, Jennifer Cole, Mark Hasegawa-Johnson, and Margaret Fleck. Feature sets for the automatic detection of prosodic prominence. In *New Tools and Methods for Very Large Scale Phonetics Research*. University of Pennsylvania, Jan. 2011.

[169] Hosung Nam, Vikramjit Mitra, Mark Tiede, Mark Hasegawa-Johnson, Carol Espy-Wilson, Elliot Saltzman, and Louis Goldstein. Automatic gestural annotation of the u. wisconsin x-ray microbeam corpus. In *Workshop on New Tools and Methods for Very Large Scale Phonetics Research*. University of Pennsylvania, Jan. 2011.

[170] Tim Mahrt, Jui-Ting Huang, Yoonsook Mo, Margaret Fleck, Mark Hasegawa-Johnson, and Jennifer Cole. Optimal models of prosodic prominence using the bayesian information criterion. In *Proc. Interspeech*, pages 2037–2040, 2011.

[171] R. Mertens, P.-S. Huang, L. Gottlieb, G. Friedland, and A. Divakaran. On the application of speaker diarization to audio concept detection for multimedia retrieval. In *IEEE International Symposium on Multimedia*, pages 446–451, 2011.

[172] Suma Bhat, Mark Hasegawa-Johnson, and Richard Sproat. Automatic fluency assessment by signal-level measurement of spontaneous speech. In *INTERSPEECH Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*, 2010.

[173] Suma Bhat, Richard Sproat, Mark Hasegawa-Johnson, and Fred Davidson. Automatic fluency assessment using thin-slices of spontaneous speech. In *Language Testing Research Colloquium*, Denver, CO, 2010.

[174] Lae-Hoon Kim, Mark Hasegawa-Johnson, Gerasimos Potamianos, and Vit Libal. Joint estimation of doa and speech based on em beamforming. In *ICASSP*, 2010.

[175] Hao Tang, Mark Hasegawa-Johnson, and Thomas S. Huang. Non-frontal view facial expression recognition. In *ICME*, pages 1202–1207, 2010.

[176] Su-Youn Yoon, Mark Hasegawa-Johnson, and Richard Sproat. Landmark-based automated pronunciation error detection. In *Proceedings of Interspeech*, pages 614–617, 2010.

[177] I. Yücel Özbek, Mark Hasegawa-Johnson, and Mübeccel Demirekler. Formant trajectories for acoustic-to-articulatory inversion. In *Proc. Interspeech*, pages 2807–2810, Brighton, September 2009.

[178] Su-Youn Yoon, Mark Hasegawa-Johnson, and Richard Sproat. Automated pronunciation scoring using confidence scoring and landmark-based svm. In *Proc. Interspeech*, pages 1903–1906, Brighton, September 2009.

[179] Bowon Lee and Mark Hasegawa-Johnson. A phonemic restoration approach for automatic speech recognition with highly nonstationary background noise. In *DSP in Cars workshop*, Dallas, July 2009.

[180] Su-Youn Yoon, Mark Hasegawa-Johnson, and Richard Sproat. Automated pronunciation scoring for l2 english learners. In *CALICO workshop*, 2009.

[181] Xiaodan Zhuang, Jing Huang, Gerasimos Potamianos, and Mark Hasegawa-Johnson. Acoustic fall detection using gaussian mixture models and gmm supervectors. In *ICASSP*, pages 69–72, 2009.

[182] Lae-Hoon Kim and Mark Hasegawa-Johnson. Optimal speech estimator considering room response as well as additive noise: Different approaches in low and high frequency range. In *ICASSP*, pages 4573–4576, 2008.

[183] Bryce Lobdell, Mark Hasegawa-Johnson, and Jont B. Allen. Human speech perception and feature extraction. In *Proc. Interspeech*, pages 1797–1800, 2008.

[184] Yoonsook Mo, Jennifer Cole, and Mark Hasegawa-Johnson. Frequency and repetition effects outweigh phonetic detail in prominence perception. In *LabPhon 11*, pages 29–30, 2008.

[185] Hao Tang, Yuxiao Hu, Yun Fu, Mark Hasegawa-Johnson, and Thomas S. Huang. Real-time conversion from a single 2d face image to a 3d text-driven emotive audio-visual avatar. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1205–1208, 2008.

[186] Shuicheng Yan, Xi Zhou, Ming Liu, Mark Hasegawa-Johnson, and Thomas S. Huang. Regression from patch kernel. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[187] Ming Liu, Xi Zhou, Mark Hasegawa-Johnson, Thomas S. Huang, and Zhengyou Zhang. Frequency domain correspondence for speaker normalization. In *Proc. Interspeech*, pages 274–277, Antwerp, August 2007.

[188] Karen Livescu, Ozgur Cetin, Mark Hasegawa-Johnson, Simon King, Chris Bartels, Nash Borges, Arthur Kantor, Partha Lal, Lisa Yung, Ari Bezman, Stephen Dawson-Haggerty, Bronwyn Woods, Joe Frankel, Matthew Magimai-Doss, and Kate Saenko. Articulatory feature-based methods for acoustic and audio-visual speech recognition: Summary from the 2006 jhu summer workshop. In *ICASSP*, pages 621–624, May 2007.

[189] Mark Hasegawa-Johnson. Multi-stream approach to audiovisual automatic speech recognition. In *IEEE 9th Workshop on Multimedia Signal Processing (MMSP)*, pages 328–331, 2007.

[190] Weimo Zhu, Mark Hasegawa-Johnson, Karen Chapman-Novakofski, and Arthur Kantor. Cellphone-based nutrition e-diary. In *National Nutrient Database Conference*, 2007.

[191] Weimo Zhu, Mark Hasegawa-Johnson, Arthur Kantor, Dan Roth, Yong Gao, Youngsik Park, and Lin Yang. E-coder for automatic scoring physical activity diary data: Development and validation. In *ACSM*, 2007.

[192] Rahul Chitturi and Mark Hasegawa-Johnson. Novel time-domain multi-class svms for landmark detection. In *Proc. Interspeech*, pages paper 1904–Thu1CaP.14, September 2006.

[193] Lae-Hoon Kim, Mark Hasegawa-Johnson, and Keung-Mo Sung. Generalized optimal multi-microphone speech enhancement using sequential minimum variance distortionless response (MVDR) beamforming and postfiltering. In *ICASSP*, pages III:65–8, May 2006.

[194] Rahul Chitturi and Mark Hasegawa-Johnson. Novel entropy based moving average refiners for HMM landmarks. In *Proc. Interspeech 2006*, pages paper 1911–Wed1FoP.8, 2006.

[195] Rajiv Reddy and Mark Hasegawa-Johnson. Analysis of pitch contours in repetition-disfluency using stem-ml. In *Midwest Computational Linguistics Colloquium*, 2006.

[196] Lae-Hoon Kim and Mark Hasegawa-Johnson. Generalized multi-microphone spectral amplitude estimation based on correlated noise model. In *119th Convention of the Audio Engineering Society*, New York, October 2005.

[197] Weimo Zhu, Mark Hasegawa-Johnson, and Mital Arun Gandhi. Accuracy of voice-recognition technology in collecting behavior diary data. In *Association of Test Publishers (ATP): Innovations in Testing*, March 2005.

[198] Jennifer Cole, Mark Hasegawa-Johnson, Chilin Shih, Eun-Kyung Lee, Heejin Kim, H. Lu, Yoonsook Mo, and Tae-Jin Yoon. Prosodic parallelism as a cue to repetition and hesitation disfluency. In *Disfluency In Spontaneous Speech (DISS'05)*, pages 53–58, Aix-en-Provence, France, 2005.

[199] Arthur Kantor, Weimo Zhu, and Mark Hasegawa-Johnson. Restricted domain speech classification using automatic transcription and svms. In *Midwest Computational Linguistics Colloquium*, 2005.

[200] Soo-Eun Chang, Nicoline Ambrose, and Mark Hasegawa-Johnson. An mri (dti) study on children with persistent developmental stuttering. In *ASHA Convention*, November 2004.

[201] Mital Gandhi and Mark Hasegawa-Johnson. Source separation using particle filters. In *Proc. Interspeech*, pages 2673–2676, October 2004.

[202] Tae-Jin Yoon, Sandra Chavarria, Jennifer Cole, and Mark Hasegawa-Johnson. Intertranscriber reliability of prosodic labeling on telephone conversation using tobi. In *Proc. Interspeech*, pages 2729–2732, October 2004.

[203] Sandra Chavarria, Taejin Yoon, Jennifer Cole, and Mark Hasegawa-Johnson. Acoustic differentiation of ip and ip boundary levels: Comparison of l- and l-l% in the switchboard corpus. In *Speech Prosody*, pages 333–336, Nara, Japan, March 2004.

[204] Heejin Kim, Jennifer Cole, Hansook Choi, and Mark Hasegawa-Johnson. The effect of accent on acoustic cues to stop voicing and place of articulation in radio news speech. In *SpeechProsody*, pages 29–32, Nara, Japan, March 2004.

[205] Mark Hasegawa-Johnson, Stephen Levinson, and Tong Zhang. Automatic detection of contrast for speech understanding. In *Proc. Interspeech 2004*, pages 581–584, 2004.

[206] Tae-Jin Yoon, Heejin Kim, and Sandra Chavarría. Local acoustic cues distinguishing two levels of prosodic phrasing: Speech corpus evidence. In *Labphon 9*, University of Illinois at Urbana-Champaign, 2004.

[207] Ken Chen, Mark Hasegawa-Johnson, Aaron Cohen, Sarah Borys, and Jennifer Cole. Prosody dependent speech recognition with explicit duration modelling at intonational phrase boundaries. In *Proc. Interspeech*, pages 393–396, September 2003.

[208] Ken Chen and Mark Hasegawa-Johnson. Improving the robustness of prosody dependent language modeling based on prosody syntax cross-correlation. In *ASRU*, 2003.

[209] Ken Chen, Mark Hasegawa-Johnson, and Jennifer Cole. Prosody dependent speech recognition on radio news. In *IEEE Workshop on Statistical Signal Processing*, St. Louis, MO, 2003.

[210] Ken Chen, Mark Hasegawa-Johnson, and Sung-Suk Kim. An intonational phrase boundary and pitch accent dependent speech recognizer. In *International Conference on Systems, Cybernetics, and Intelligence*, 2003.

[211] Jennifer Cole, Hansook Choi, Heejin Kim, and Mark Hasegawa-Johnson. The effect of accent on the acoustic cues to stop voicing in radio news speech. In *ICPhS*, pages 2665–2668, 2003.

[212] Mohamed Kamal Omar and Mark Hasegawa-Johnson. Non-linear maximum likelihood feature transformation for speech recognition. In *Proc. 8th European Conference on Speech Communication and Technology (Eurospeech 2003)*, pages 2497–2500, 2003.

[213] Mohammed Kamal Omar, Ken Chen, Mark Hasegawa-Johnson, and Yigal Brandman. An evaluation of using mutual information for selection of acoustic-features representation of phonemes for speech recognition. In *Proc. Interspeech*, pages 2129–2132, Denver, CO, September 2002.

[214] Mohammed Kamal Omar and Mark Hasegawa-Johnson. Maximum mutual information based acoustic features representation of phonological features for speech recognition. In *ICASSP*, pages I:81–84, May 2002.

[215] Wira Gunawan and Mark Hasegawa-Johnson. Plp coefficients can be quantized at 400 bps. In *ICASSP*, pages 2.2.1–4, Salt Lake City, UT, 2001.

[216] Mark Hasegawa-Johnson. Multivariate-state hidden markov models for simultaneous transcription of phones and formants. In *ICASSP*, pages 1323–1326, Istanbul, 2000.

[217] Mark Hasegawa-Johnson. Time-frequency distribution of partial phonetic information measured using mutual information. In *Proc. 6th International Conference on Spoken Language Processing (ICSLP 2000)*, pages vol. 4, 133–136, 2000.

[218] Jun Huang, Stephen Levinson, and Mark Hasegawa-Johnson. Signal approximation in Hilbert space and its application on articulatory speech synthesis. In *Proc. 6th International Conference on Spoken Language Processing (ICSLP 2000)*, pages vol. 2, 775–778, 2000.

[219] Mark Johnson. Automatic context-sensitive measurement of the acoustic correlates of distinctive features. In *ICSLP*, pages 1639–1643, Yokohama, 1994.

[220] Mark A. Johnson and Tomohiko Taniguchi. Low-complexity multi-mode vxc using multi-stage optimization and mode selection. In *ICASSP*, pages 221–224, Toronto, Canada, 1991.

[221] Tomohiko Taniguchi, Fumio Amano, and Mark A. Johnson. Improving the performance of celp-based speech coding at low bit rates. In *International Symposium on Circuits and Systems*, Singapore, 1991.

[222] Tomohiko Taniguchi, Mark A. Johnson, and Yasuji Ohta. Pitch sharpening for perceptually improved celp, and the sparse-delta codebook for reduced computation. In *ICASSP*, pages 241–244, Toronto, Canada, 1991.

[223] Mark A. Johnson and Tomohiko Taniguchi. Computational reduction in sparse-codebook celp using backward-weighting of the input. In *Institute of Electr. and Information and Comm. Eng. Symposium DSP 90-15*, pages 61–66, Hakata, 1990.

[224] Mark A. Johnson and Tomohiko Taniguchi. Pitch-orthogonal code-excited LPC. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 542–546, San Diego, CA, 1990.

[225] Tomohiko Taniguchi, Mark A. Johnson, and Yasuji Ohta. Multi-vector pitch-orthogonal LPC: quality speech with low complexity at rates between 4 and 8 kbps. In *ICSLP*, pages 113–116, Kobe, 1990.

Student Ph.D. Theses

[1] Mahir Morshed. *Direct Exposure of Linguistic Principles in Model Architectures for Downstream Processing*. PhD thesis, University of Illinois, 5 2025.

[2] John Harvill. *Speech classification and lexical semantic modeling via self-supervision and knowledge transfer*. PhD thesis, University of Illinois, April 2024.

[3] Heting Gao. *Unsupervised speech technology for low-resource languages*. PhD thesis, University of Illinois, March 2024.

[4] Jialu Li. *Breaking down barriers: advancing interdisciplinary speech applications in early children's development*. PhD thesis, University of Illinois, March 2024.

[5] Zhonghao Wang. *Reasoning, scaling, generating with vision-language models*. PhD thesis, University of Illinois, March 2024.

[6] Oshane Odane Thomas. *Evolution of primate cuboid form and the application of learning methods to the analysis of morphological and behavioral phenotypes*. PhD thesis, University of Illinois, June 2023.

[7] Liming Wang. *Multimodal spoken unit discovery with paired and unpaired modalities.* PhD thesis, University of Illinois, May 2023.

[8] Ali Abavisani. *The role of talker change, vowel change and speech perceptual measure in hearing impaired phone recognition.* PhD thesis, University of Illinois, May 2022.

[9] Yuchen Fan. *Sparse representation in deep vision models.* PhD thesis, University of Illinois, November 2021.

[10] Yuqian Zhou. *Towards practical deep learning based image restoration model.* PhD thesis, University of Illinois, November 2021.

[11] Leda Sarı. *Learning speech embeddings for speaker adaptation and speech understanding.* PhD thesis, University of Illinois, April 2021.

[12] Kaizhi Qian. *Deep generative models for speech editing.* PhD thesis, University of Illinois, November 2020.

[13] Wenda Chen. *Modeling phones, keywords, topics and intents in spoken languages.* PhD thesis, University of Illinois, July 2019.

[14] Xuesong Yang. *Dealing with linguistic mismatches for automatic speech recognition.* PhD thesis, University of Illinois, May 2019.

[15] Amit Das. *Speech Recognition with Probabilistic Transcriptions and End-to-End Systems Using Deep Learning.* PhD thesis, University of Illinois, 2018.

[16] Mary Pietrowicz. *Exposing the Hidden Vocal Channel: Analysis of Vocal Expression.* PhD thesis, University of Illinois, 2017.

[17] Roger Serwy. *Hilbert Phase Methods for Glottal Activity Detection.* PhD thesis, University of Illinois, 2017.

[18] Yang Zhang. *Application of Generative Models in Speech Processing Tasks.* PhD thesis, University of Illinois, 2017.

[19] Sujeeth Bharadwaj. *A theory of (almost) zero resource speech recognition.* PhD thesis, University of Illinois, June 2015.

[20] Po-Sen Huang. *Shallow and deep learning for audio and natural language processing.* PhD thesis, University of Illinois, January 2015.

[21] Mahmoud Abunasser. *Computational Measures of Linguistic Variation: A Study of Arabic Varieties.* PhD thesis, University of Illinois, 2015.

[22] Jui-Ting Huang. *Semi-Supervised Learning for Acoustic and Prosodic Modeling in Speech Applications.* PhD thesis, University of Illinois, 2012.

[23] Harsh Vardhan Sharma. *Acoustic Model Adaptation for Recognition of Dysarthric Speech.* PhD thesis, University of Illinois, 2012.

[24] Xiaodan Zhuang. *Modeling Audio and Visual Cues for Real-world Event Detection.* PhD thesis, University of Illinois, April 2011.

[25] Boon Pang Lim. *Computational Differences between Whispered and Non-whispered Speech.* PhD thesis, University of Illinois, 2011.

[26] Lae-Hoon Kim. *Statistical Model Based Multi-Microphone Speech Processing: Toward Overcoming Mismatch Problem.* PhD thesis, University of Illinois, August 2010.

[27] Arthur Kantor. *Pronunciation modeling for large vocabulary speech recognition.* PhD thesis, University of Illinois, 2010.

[28] Sarah Borys. *Lovable Indestructible Grad Student of Chaos.* PhD thesis, University of Illinois, 2009.

[29] Bryce Lobdell. *Models of Human Phone Transcription in Noise Based on Intelligibility Predictors.* PhD thesis, University of Illinois, 2009.

[30] Taejin Yoon. *A Predictive Model of Prosody Through Grammatical Interface: A Computational Approach.* PhD thesis, University of Illinois, 2007.

[31] Bowon Lee. *Robust Speech Recognition in a Car Using a Microphone Array.* PhD thesis, University of Illinois, 2006.

[32] Yanli Zheng. *Acoustic Modeling and Feature Extraction for Speech Recognition.* PhD thesis, University of Illinois, 2005.

[33] Ken Chen. *Prosody Dependent Speech Recognition on American Radio News Speech.* PhD thesis, University of Illinois, 2004.

[34] Mohamed Kamal Mahmoud Omar. *Acoustic Feature Design for Speech Recognition: A Statistical Information-Theoretic Approach.* PhD thesis, University of Illinois, 2003.

[35] Mark A. Hasegawa-Johnson. *Formant and Burst Spectral Measurements with Quantitative Error Models for Speech Sound Classification.* PhD thesis, MIT, 1996.

Student M.S. Theses

[1] Kai Chieh Chang. Fusing multimodal neural networks: a study on sleep classification and sound event localization and detection. Master's thesis, University of Illinois, 2024.

[2] Feiyu Zhang. End-to-end modeling for code-switching automatic speech recognition. Master's thesis, University of Illinois, 2024.

[3] Seunghyun Lee. Explainable artificial intelligence for inclusive automatic speech recognition. Master's thesis, University of Illinois, 2023.

[4] Chak Ho Chan. Speechsplit2: Disentangling speech information streams without exhaustive bottleneck fine-tuning. Master's thesis, University of Illinois, 2022.

[5] Haozhong Guan. Study on speech emotion recognition based on deep learning. Master's thesis, University of Illinois, 2022.

[6] John Harvill. Regularization for dysarthric speech recognition and telemedicine applications. Master's thesis, University of Illinois, 2022.

[7] Mahir Morshed. Cross-lingual articulation information transfer with progressive networks for speech recognition. Master's thesis, University of Illinois, 2022.

[8] Heting Gao. Improving multilingual speech recognition systems. Master's thesis, University of Illinois, 2021.

[9] Junrui Ni. Enforcing constraints for multi-lingual and cross-lingual speech-to-text systems. Master's thesis, University of Illinois, 2021.

[10] Kiran Ramnath. Thumbnail for fact-based visual question answering using knowledge graph embeddings fact-based visual question answering using knowledge graph embeddings. Master's thesis, University of Illinois, 2021.

[11] Liming Wang. A translation framework for discovering word-like units from visual scenes and spoken descriptions. Master's thesis, University of Illinois, 2020.

[12] Ningkai Wu. Semi-supervised cycle-consistency training for end-to-end asr using unpaired speech. Master's thesis, University of Illinois, 2020.

[13] Junzhe Zhu. Multi-decoder DPRNN high accuracy source counting and separation. Master's thesis, University of Illinois, 2020.

[14] Yijia Xu. Acoustic event, spoken keyword and emotional outburst detection. Master's thesis, University of Illinois, 2019.

[15] Kaizhi Qian. Speech enhancement using deep dilated CNN. Master's thesis, University of Illinois, 2018.

[16] Tianyilin Zhu. Thumbnail for lipreading with convolutional and recurrent neural network models lipreading with convolutional and recurrent neural network models. Master's thesis, University of Illinois, 2017.

[17] Xuesong Yang. Machine learning approaches to improving mispronunciation detection on an imbalanced corpus. Master's thesis, University of Illinois, 2015.

[18] Yang Zhang. Probabilistic generative modeling of speech. Master's thesis, University of Illinois, 2015.

[19] Austin Chen. Automatic classification of electronic music and speech/music audio content. Master's thesis, University of Illinois, 2014.

[20] Daniel Soberal. Face recognition using hidden markov model supervectors. Master's thesis, University of Illinois, 2014.

[21] Sujeeth Bharadwaj. Multiview feature learning for speech recognition. Master's thesis, University of Illinois, 2011.

[22] Po-Sen Huang. Non-speech acoustic event detection using multimodal information. Master's thesis, University of Illinois, 2011.

[23] Jeremy Tidemann. Characterization of the head-related transfer function using chirp and maximum length sequence excitation signals. Master's thesis, University of Illinois, 2011.

[24] Chi Hu. Fsm-based pronunciation modeling using articulatory phonological code. Master's thesis, University of Illinois, 2010.

[25] Sarah Borys. An svm front end landmark speech recognition system. Master's thesis, University of Illinois, 2008.

[26] Yang Li. Incremental training and growth of artificial neural networks. Master's thesis, University of Illinois, 2008.

[27] Harsh Vardhan Sharma. Universal access: Experiments in automatic recognition of dysarthric speech. Master's thesis, University of Illinois, 2008.

[28] Christopher Co. Room reconstruction and navigation using acoustically obtained room impulse responses and a mobile robot platform. Master's thesis, University of Illinois, 2005.

[29] Aaron Cohen. A survey of machine learning methods for predicting prosody in radio speech. Master's thesis, University of Illinois, 2004.

[30] Zhinian Jing. Voice index and frame index for recognition of digits in speech background. Master's thesis, University of Illinois, 2002.

[31] Jul Setsu Cha. Articulatory speech synthesis of female and male talkers. Master's thesis, UCLA, 2000.

[32] Wira Gunawan. Distributed speech recognition. Master's thesis, University of Illinois, 2000.

Student B.S. Thesis, Unpublished Manuscripts and Talks

[1] Giang Le, Yinglun Sun, Maliha Jahan, Helin Wang, Thomas Thebaud, Zsuzsanna Fagyal, Mark Hasegawa-Johnson, Laureano Moro-Velazquez, and Najim Dehak. "you are what, chinese?" modeling identifications of race, ethnicity, and national origin as stancetaking in a large podcast corpus of american english. presented at NWAV New Ways of Analyzing Variation, 11 2024.

[2] Mark Hasegawa-Johnson, Heejin Kim, Xiuwen Zheng, Meg Dickinson, Erik Hege, Clarion Mendes, Chris Zwilling, Shadi Abou-Zahra, Leah Findlater, Aaron Gustafson, Bob MacDonald, Mike Shebanek, Adina Bradshaw, Shawnise Carter, Blair Casey, Marie Moore Channell, Kayla Ferguson, Heather Hodges, Laura Mattie, and Lorraine Ramig. Community-supported shared infrastructure in support of speech accessibility. Unpublished presentation at the ASHA 2023 Research Symposium, 2023.

[3] Mahir Morshed. Spiral representations in end-to-end bengali articulatory feature identification. Unpublished B.S. Thesis, 2019.

[4] Mark Hasegawa-Johnson. Unwritten languages as a test case for the theory of phonetic universals. Plenary talk delivered at the International Symposium on Chinese Spoken Language Processing, 2018.

[5] Liming Wang. Multimodal semantic learning with context-correlated speeches and images. Unpublished B.S. thesis, 2018.

[6] Emmanuel Dupoux, Odette Scharenborg, Graham Neubig, Laurent Besacier, Mark Hasegawa-Johnson, Alan Black, Florian Metze, and Sebastian Stüker. The speaking rosetta stone - discovering grounded linguistic units for languages without orthography. Final report of the WS17 Frederick Jelinek Memorial Summer Workshop team, 2017.

[7] Mark Hasegawa-Johnson. Limited data settings. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.

[8] Mark Hasegawa-Johnson. Multivariate-state models for speech recognition. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.

[9] Mark Hasegawa-Johnson. Prosody. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.

[10] Mark Hasegawa-Johnson. Speech production, speech perception, and phonology. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.

[11] Mark Hasegawa-Johnson, Ed Lalor, KC LEe, Preethi Jyothi, Majid Mirbagheri, Amit Das, Giovannie Di Liberto, Brad Ekin, Chunxi Liu, Vimal Manohar, Hao Tang, Paul Hager, Tyler Kekona, and Rose Sloan. Probabilistic transcription. WS15 Group Final Presentation, 2015.

[12] Jonathan Jones. English-spanish language transfer for automatic speech recognition. Unpublished B.S. Thesis, 2014.

[13] Kaizhi Qian. Regularized estimation of gaussian mixture models for svm based speaker recognition. B.S. Thesis, University of Illinois, 2014.

[14] Raymond Yeh. Divergence guided two beams viterbi algorithm on factorial hmms. B.S. Thesis, University of Illinois, 2014.

[15] Mark Hasegawa-Johnson, Xiaodan Zhuang, Xi Zhou, Camille Goudeseune, Hao Tang, Kai-Hsiang Lin, Mohamed Omar, and Thomas Huang. Toward better real-world acoustic event detection. Unpublished presentation given at Seoul National University, May 2012.

[16] Rania Al-Sabbagh, Roxana Girju, Mark Hasegawa-Johnson, Elabbas Benmamoun, Rehab Duwairi, and Eiman Mustafawi. Using web mining techniques to build a multi-dialect lexicon of arabic. Talk delivered at the Linguistics in the Gulf Conference, 2011.

[17] David Harwath. Phonetic landmark detection for automatic language identification. Unpublished B.S. Thesis, 2010.

[18] Sujeeth Bharadwaj. Advances in sparse classification. Unpublished B.S. Thesis, 2009.

[19] David Cohen, Camille Goudeseune, and Mark Hasegawa-Johnson. Efficient simultaneous multi-scale computation of ffts. Georgia Tech publication GT-FODAVA-09-01, 2009.

[20] Sarah Borys. Lovable indestructible grad student of chaos. Cartoons published online, 2008.

[21] Karen Livescu, Özgür Çetin, Mark Hasegawa-Johnson, Simon King, Chris Bartels, Nash Borges, Arthur Kantor, Partha Lal, Lisa Yung, Ari Bezman, Stephen Dawson-Hagerty, Bronwyn Woods, Joe Frankel, Mathew Magimai-Doss, and Kate Saenko. Articulatory-feature-based methods for acoustic and audio-visual speech recognition: 2006 jhu summer workshop final report. Final report of the WS06 Johns Hopkins Summer Workshop team, 2007.

[22] Rajiv Reddy. Analysis of pitch contours in repetition-disfluency using stem-ml. 2006.

[23] Mark Hasegawa-Johnson, James Baker, Steven Greenberg, Katrin Kirchhoff, Jennifer Muller, Kemal Sonmez, Sarah Borys, Ken Chen, Amit Juneja, Katrin Kirchhoff, Karen Livescu, Srividya Mohan, Emily Coogan, and Tianyu Wang. Landmark-based speech recognition: Report of the 2004 johns hopkins summer workshop. Final report of the WS04 Johns Hopkins Summer Workshop team, 2005.

[24] Stefan Geirhofer. Feature reduction with linear discriminant analysis and its performance on phoneme recognition. Undergraduate research project, 2004.

[25] Sarah Borys. Recognition of prosodic factors and detection of landmarks for improvements to continuous speech recognition systems. B.S. Thesis, University of Illinois, 2003.

[26] Ameya Deoras. A factorial hmm approach to robust isolated digit recognition in non-stationary noise. B.S. Thesis, University of Illinois, 2003.

[27] David Petruncio. Evaluation of various features for music genre classification with hidden markov models. B.S. Thesis, University of Illinois, 2002.