

Mark Hasegawa-Johnson

Professor, Department 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

September 13, 2023

Education

- **Post-Doctoral Fellow**, 1996-9, University of California at Los Angeles. Adviser: Abeer Alwan, Electrical Engineering. NRSA Title: Factor Analysis of MRI-Derived Articulator Shapes
- **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. Electrical Engineering and Computer Science**, June 1989, Massachusetts Institute of Technology. Adviser: Jae S. Lim. Thesis: Echo Cancellation in the GSM Cellular Network

Appointments

- 2020-present: **William L. Everitt Faculty Scholar in 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.
- 2016-present: **Professor**, Informatics, **Kyoto College of Graduate Studies in Informatics**, Kyoto, Japan
- 2022: **Visiting Professor**, Kyoto University, Kyoto, Japan
- 2011-2020: **Professor**, Electrical and Computer Engineering, University of Illinois
- 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
- 2005-2011: **Associate Professor**, Electrical and Computer Engineering, University of Illinois
- 1999-2005: **Assistant Professor**, Electrical and Computer Engineering, University of Illinois
- 1996-1999: **Post-Doctoral Fellow, 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

Professional Society Offices and Conference General Management

1. **Treasurer, ISCA** (International Speech Communication Association, 2013-2021)
2. **Liaison**, Special Interest Group on Machine Learning (SIGML) of the International Speech Communication Association (ISCA) (2010-2020)
3. **Secretary**, Speech Prosody Special Interest Group (SProSIG) of the International Speech Communication Association (ISCA) (2010-2018)
4. **Conference General Chair**, Fifth International Conference on Speech Prosody (2010)

Editorships and Conference Technical Management

1. **Senior Area Editor**, IEEE Transactions on Audio, Speech and Language, 2018-2023
2. **Technical Program Co-Chair**, Interspeech 2022
3. **Member, Speech and Language Technical Committee (SLTC)**, IEEE Signal Processing Society (2011-2017)
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).

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

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)

Awards

1. **Outstanding Reviewer**, ICASSP 2023
2. **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.
3. **Second Place**, Diagnosing CoVID-19 from Acoustics (DiCOVA) challenge competition, ICASSP 2022
4. **Third Place**, Diagnosing CoVID-19 from Acoustics (DiCOVA) challenge competition, Interspeech 2021
5. **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,” <https://ieeexplore.ieee.org/document/7194774>.
6. **William L. Everitt Faculty Scholar Award**, ECE Department, University of Illinois, 2020-2021
7. University of Illinois **List of Teachers Rated “Excellent”** by their Students, ten times including Fall 2019, Spring 2020, Fall 2020, Fall 2021, Fall 2022. [https://citl.illinois.edu/citl-101/measurement-evaluation/teaching-evaluation/teaching-evaluations-\(ices\)/teachers-ranked-as-excellent](https://citl.illinois.edu/citl-101/measurement-evaluation/teaching-evaluation/teaching-evaluations-(ices)/teachers-ranked-as-excellent)
8. **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
9. 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.

10. **Dean’s Award for Excellence in Research**, University of Illinois College of Engineering, 2012
11. **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
12. **Third Place**, Star Challenge Multimedia Information Retrieval Competition, A*STAR 2008
13. **First Place**, Acoustic Event Recognition competition, 2007 CLEAR (Classification of Events, Activities and Relationships) Evaluation and Workshop (Stiefelhaven et al., “The CLEAR 2007 Evaluation,” LNCS 4625:3-34, 2008)
14. **Outstanding Advisers List**, University of Illinois College of Engineering, April 2006
15. **Best Reviewer**, Neural Information Processing Systems (NIPS), 2005
16. **Honorary Initiate**, Alpha Chapter of Eta Kappa Nu (Electrical and Computer Engineering Honor Society), 2003
17. **Frederick V. Hunt Post-Doctoral Research Fellow**, Acoustical Society of America, 1996
18. **Charles LeGeyt Fortescue Graduate Fellow**, IEEE, 1989
19. Eta Kappa Nu, Tau Beta Pi, Sigma Xi, Phi Beta Kappa

Conference and Workshop Keynotes and Invited Talks

1. “Grapheme-to-Phoneme Transduction for Cross-Language ASR,” Workshop on Statistical Language and Speech Processing, 2021.
2. “Multimodal Distant Supervision,” NIPS workshop on Self-Supervised Learning for Speech and Audio Processing, 2020
3. “Distant Supervision for Cross-Language Speech Adaptation.” Keynote speech, ASRU 2019 (IEEE Workshop on Automatic Speech Recognition and Understanding)
4. “Multimedia AI and User Trust,” Invited talk, AI Association of Korea Workshop on Artificial Intelligence and Blockchain, Seoul, 11/2019
5. “Unwritten Languages as a Test Case for the Theory of Phonetic Universals.” Keynote speech, ISCSLP 2018 (International Symposium on Chinese Spoken Language Processing), 11/2018
6. “Prosody in Speech Technology.” Series of four plenary lectures, ISCA Winter School on Speech and Audio Processing, 2016.

Workshop and Panel Organization Activities

1. **Team Member**, Jelinek Speech and Language Technology (JSALT) Workshop WS17, “The Speaking Rosetta Stone - Discovering Grounded Linguistic Units for Languages without Orthography,” Pittsburgh, PA, 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, WA, June–August, 2015
3. **Panel Organizer**, VAC Consortium Working Group on Multimedia Analytics (Adelphi, MD; May 2011)
4. **Workshop Co-Chair**, SPREI Speech Production Workshop (Urbana, IL; May 2011)
5. **Workshop Co-Chair**, Illinois Speech Day (Chicago, IL; May 2009, May 2010, May 2011)
6. **Workshops Co-Chair**, HLT/NAACL 2009
7. **Team Member**, DARPA/NSF CLSP Summer Research Workshop, Articulatory-Feature Based Speech Recognition, Baltimore, MD, June–August, 2006
8. **Team Leader**, DARPA/NSF CLSP Summer Research Workshop, Landmark-Based Speech Recognition, Baltimore, MD, June–August, 2004

9. **Technical Committee or Conference Reviewer:** AAAI; ACL; AISTATS; UIUC Allerton; AP-SIPA; IEEE ASRU; CBMI; EMNLP; ACL HLT; IEEE ICASSP; IEEE ICC; ICML; ICPR; ICPPIT; ICPRAM; ISCA Interspeech; LabPhon; MCLC; NAACL; NeurIPS (Voted ‘Best Reviewer:’ 2005, 2020); ISCA/ACL SLPAT; ISCA SLaTE; ISCA SPASR; UIUC SPREI; IEEE SLT
10. **Proposal Reviewer:** National Science Foundation (NSF), National Institutes of Health (NIH), Netherlands Organization for Scientific Research (NWO), National Science and Engineering Research Council of Canada (NSERC), Qatar National Research Fund (QNRF), Springer Academic Publishing

Course Director

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

Students and Collaborators

- **PhD Students (Graduated):** Mohamed Kamal Omar (12/2003; IBM), Ken Chen (5/2004; University of Texas MD Anderson, Bioinformatics and Computational Biology), Yanli Zheng (12/2004; FICO), Bowon Lee (12/2006; Inha University), Bryce Lobdell (5/2009; openbi.com), Lae-Hoon Kim (8/2010; Qualcomm), Arthur Kantor (10/2010; IBM), Boon Pang Lim (12/2010; Novumind), Xiaodan Zhuang (5/2011; Apple), Andreas Ehmann (12/2011; Pandora), Jui-Ting Huang (1/2012; Facebook); Harsh Vardhan Sharma (2/2012; Credit Karma); Sujeeth Bharadwaj (5/2015; Microsoft); Po-Sen Huang (5/2015; Microsoft); Roger Serwy (5/2017; Enthought); Yang Zhang (5/2017; MIT IBM Watson AI Laboratory); Mary Pietrowicz (12/2017; IBM); Amit Das (8/2018; Microsoft Research); Xuesong Yang (8/2018; Kwai AI); Wenda Chen (2019; Intel Labs); Kaizhi Qian (12/2020; MIT IBM Watson AI Laboratory); Leda Sari (5/2021; Facebook AI Research); Yuchen Fan (5/2022; Meta Reality Labs); Yuqian Zhou (5/2022; Adobe); Ali Abavisani (5/2022; Bosch); Liming Wang (5/2023; MIT)
- **Post-Doctoral Fellows:** Jeung-Yoon Choi (2002-4; Yonsei University); Heejin Kim (2006-10; University of Illinois); Kyung-Tae Kim (2008-10; Samsung); Arthur Kantor (2010-11; IBM); Suma Bhat (2011-4; UIUC); Preethi Jyothi (2013-6; IIT Bombay)
- **Visiting Professors and Visiting Scholars:** Sung-Suk Kim (Yong-In University; 2002-2003), Sung-Tae Jung (Wong-Kwang University; 2004-2005), Yanxiang Chen (University of Science and Technology of China; 2005-6), Zhijian Ou (Tsinghua University; 2014-5), Yanlu Xie (Beijing Language and Culture University; 2015-6)
- **Post-Graduate and Post-Doctoral Advisors:** Jae S. Lim (MIT), Kenneth N. Stevens (MIT), Abeer Alwan (UCLA)

Grants Received

1. Factor Analysis of MRI-Derived Articulator Shapes. NIH Individual National Research Service Award, 1999.
2. PI: Factor Analysis of the Tongue Shapes of Speech. University of Illinois Research Board, 1999-2000.
3. PI: Immersive Headphone-free Virtual Reality Audio. University of Illinois Research Board, 2001-2002.
4. PI: Prosody-Dependent Speech Recognition. University of Illinois Critical Research Initiative, 2002-2004.
5. PI: CAREER: Landmark-Based Speech Recognition in Music and Speech Backgrounds. NSF IIS 01-32900, 2002-2007.
6. PI: Acoustic Features for Phoneme Recognition. Phonetact Incorporated, 2002.
7. PI: Audiovisual Speech Recognition: Data Collection and Feature Extraction in Automotive Environment. Motorola Communications Center RPS 19, 2002-2005.

8. Co-PI: Development and Validation of An E-diary System for Assessing Physical Activity and Travel Behaviors. Robert Wood Johnson Foundation, 2003-2004.
9. PI: Prosodic, Intonational, and Voice Quality Correlates of Disfluency. NSF IIS 04-14117, 2004-2007.
10. Co-PI: Automated Methods for Second-Language Fluency Assessment. University of Illinois Critical Research Initiative, 2005-2007.
11. PI: Audiovisual Distinctive-Feature-Based Recognition of Dysarthric Speech. NSF IIS 05-34106, 2006-2010.
12. PI: Description and Recognition of Audible and Visible Dysarthric Phonology, NIH, PHS 1 R21 DC008090A, 2006-2009.
13. PI: Rhythmic Organization of Durations for Automatic Speech Recognition. UIUC Research Board, 2005-6.
14. Co-PI: Cell Phone Annoyance Factors. QUALCOMM, Inc., 2005-7.
15. Co-PI: Audiovisual Emotional Speech AVATAR. Motorola Communications Center RPS 31, 2005-7.
16. Co-PI: DHB: Fluency and the Dynamics of Second Language Acquisition. NSF IIS 06-23805, 2006-10.
17. Co-PI: RI-Collaborative Research: Landmark-based robust speech recognition using prosody-guided models of speech variability. NSF IIS 07-03624, 2007-12.
18. PI: RI Medium: Audio Diarization - Towards Comprehensive Description of Audio Events. NSF IIS 08-03219, 2008-10.
19. PI: FODAVA-Partner: Visualizing Audio for Anomaly Detection. NSF CCF 08-07329, 2008-13.
20. Co-PI: Opportunistic Sensing for Object and Activity Recognition from Multi-Modal, Multi-Platform Data. ARO W911NF-09-1-0383, 2009-14.
21. PI: Multi-dialect phrase-based speech recognition and machine translation for Qatari broadcast TV. Qatar National Research Fund NPRP 09-410-1-069, 2010-3.
22. Co-PI: CDI-Type II: Collaborative Research: Groupscape: Instrumenting Research on Interaction Networks in Complex Social Contexts, NSF 0941268, 2010-4.
23. Co-PI: Speech Production Research Initiative, University of Illinois Graduate College Focal Point Program, 2010-11
24. Faculty Mentor: FY 2011 Summer Undergraduate Research Fellowship SURF NIST Gaithersburg, NIST COM 70NANB11H087, 2011
25. PI: Pseudo-intelligent mediators ("Robo-Buddies") to improve communication between students with and students without physical disabilities, Illinois Innovation Initiative (In3), 2012-4
26. Co-PI: Conversation Strategies for Students With and Students Without Physical Disabilities, University of Illinois Graduate College Focal Point Program, 2012-3
27. Co-PI: AHRQ R21-Hs022948, Collaborative Patient Portals: Computer-based Agents and Patients ' Understanding of Numeric Health Information, 2014-6
28. LPI: QNRF NPRP 7-766-1-140, The Family as the Unit of Intervention for Speech-Generating Augmentative/Assistive Communication, 2014-8
29. PI: Illinois Learning Sciences Design Initiative (ILSDI), University of Illinois. "Capturing. Transcribing. Searching. Analyzing. Adaptive: Learning in a curated classroom." 2015-6
30. Co-PI: Institute for Infocomm Research (I²R), Agency for Science, Technology, Advancement and Research (ASTAR), Singapore. "Mismatched Crowdsourcing for 80-Language Speech Recognition." 2015-7
31. PI: Advanced Digital Sciences Center (ADSC), Singapore. "Noisy Channel Models for Massively Multilingual Automatic Speech Recognition." 2015-7
32. PI: NSF. "EAGER: Matching Non-Native Transcribers to the Distinctive Features of the Language Transcribed." 2015-8
33. Co-PI: DARPA LORELEI. "LanguageNet: Transfer Learning Across a Language Similarity Networks." 2015-9
34. Co-PI: NSF IIS 19-10319. "RI: Small: Collaborative Research: Automatic Creation of New Phone Inventories." 07/2019-06/2022

35. Co-I of Subaward from Duke University: NIH 1 R34 DA050256. “HEAL Consortium: Establishing Innovative Approaches for the HEALthy Brain and Child Development Study.” 11/2019-10/2021
36. PI of Subaward from KAIST: Institute for Information & Communications Technology Promotion (IITP), Korea. “Deep F-measure Maximization for Fairness in Speech Understanding.” 2020-2021
37. Co-PI: C3.ai Digital Transformation Institute. “Adding Audio-Visual Cues to Signs and Symptoms for Triaging Suspected or Diagnosed COVID-19 Patients.” 6/2020-12/2021
38. Co-PI: Virtual Reality Visualization of Complex and Unstructured Data and Relationships, AFWERX, 1/2022-2/2022
39. PI: NSF FAI: A New Paradigm for the Evaluation and Training of Inclusive Automatic Speech Recognition, 2/2022-2/2025
40. Co-PI: Virtual Reality Visualization of Complex and Unstructured Data and Relationships, AFWERX, 10/2022-6/2023
41. PI: Nova Habilitas: Speech Accessibility Project. Amazon, Apple, Google, Meta and Microsoft, 9/2022-8/2024

Book Chapters

- [1] Andrew Rosenberg and Mark Hasegawa-Johnson. Automatic prosody labeling and assessment. In Carlos Gussenhoven and Aojun Chen, editors, *Oxford Handbook of Language Prosody*, pages 646–656. Oxford University Press, 2021.
- [2] 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.
- [3] Mark Hasegawa-Johnson and Abeer Alwan. Speech coding: Fundamentals and applications. In J. Proakis, editor, *Wiley Encyclopedia of Telecommunications and Signal Processing*. Wiley and Sons, NY, December 2002.

Conference Papers

- [1] 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 *APSIPA ASC*, 11 2023.
- [2] Wonjune Kang, Mark Hasegawa-Johnson, and Deb Roy. End-to-end zero-shot voice conversion with location-variable convolutions. In *Interspeech*, 2023.
- [3] 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 *Interspeech*, 2023.
- [4] Liming Wang, Mark Hasegawa-Johnson, and Chang D. Yoo. A theory of unsupervised speech recognition. In *ACL*, 7 2023.
- [5] 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.
- [6] 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.
- [7] 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 *Interspeech*, 2023.

- [8] Wanyue Zhai and Mark Hasegawa-Johnson. Wav2ToBI: a new approach to automatic ToBI transcription. In *Interspeech*, 2023.
- [9] 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.
- [10] 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.
- [11] John Harvill, Roxana Girju, and Mark Hasegawa-Johnson. Syn2Vec: Synset colexification graphs for lexical semantic similarity. In *Proc. NAACL*, page 5259–5270, 2022.
- [12] John Harvill, Mark Hasegawa-Johnson, and Chang D. Yoo. Frame-Level Stutter Detection. In *Proc. Interspeech 2022*, pages 2843–2847, 2022.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Leda Sari, 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.
- [17] 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.
- [18] Raymond Yeh, Mark Hasegawa-Johnson, and Alexander Schwing. Equivariance discovery by learned parameter-sharing. In *AISTATS*, 2022.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.

- [24] 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.
- [25] Kiran Ramnath, Leda Sari, 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] Junzhe Zhu, Mark Hasegawa-Johnson, and Nancy McElwain. A comparison study on infant-parent voice diarization. In *Proc. ICASSP*, pages 7178–7182, 2021.
- [31] 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.
- [32] Ali Abavisani and Mark Hasegawa-Johnson. Automatic estimation of intelligibility measure for consonants in speech. In *Proc. Interspeech*, pages 1161–1165, 2020.
- [33] Mark Hasegawa-Johnson. Multimodal distant supervision. In *NeurIPS Workshop on Self-Supervised Learning for Speech and Audio*, 2020.
- [34] Jialu Li and Mark Hasegawa-Johnson. Autosegmental neural nets: Should phones and tones be synchronous or asynchronous? In *Proc. Interspeech*, pages 1027–1031, 2020.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] Leda Sari and Mark Hasegawa-Johnson. Deep F-Measure Maximization for End-to-End Speech Understanding. In *Proc. Interspeech*, pages 1580–1584, 2020.
- [40] Leda Sari, Samuel Thomas, and Mark Hasegawa-Johnson. Training spoken language understanding systems with non-parallel speech and text. In *Proc. ICASSP*, pages 8109–8113, 2020.

- [41] Justin van der Hout, Mark Hasegawa-Johnson, and Odette Scharenborg. Evaluating automatically generated phoneme captions for images. In *Proc. Interspeech*, pages 2317–2321, 2020.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] Leda Sari, 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.
- [52] 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.
- [53] 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.
- [54] 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.
- [55] 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.
- [56] Wenda Chen, Mark Hasegawa-Johnson, and Nancy Chen. Recognizing zero-resourced languages based on mismatched machine transcriptions. In *Proc. ICASSP*, pages 5979–5983, 2018.

- [57] 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.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] Teck Yian Lim, Raymond Yeh, Yijia Xu, Minh Do, and Mark Hasegawa-Johnson. Time-frequency networks for audio super-resolution. In *Proc. ICASSP*, 2018.
- [63] 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.
- [64] 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.
- [65] 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.
- [66] 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.
- [67] 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 Merckx, 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.
- [68] 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.
- [69] 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.
- [70] Yijia Xu, Mark Hasegawa-Johnson, and Nancy L. McElwain. Infant emotional outbursts detection in infant-parent spoken interactions. In *Interspeech*, pages 242–246, 2018.
- [71] 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.
- [72] 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.

- [73] 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.
- [74] 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.
- [75] 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.
- [76] Shiyu Chang, Yang Zhang, Wei Han, Mo Yu and Xiaoxiao Guo, Wei Tan, Xiaodong Cui, Michael Witbrock, Mark Hasegawa-Johnson, and Thomas Huang. Dilated recurrent neural networks. In *NIPS*, 2017.
- [77] 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.
- [78] 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.
- [79] Amit Das, Mark Hasegawa-Johnson, and Karel Vesely. Deep autoencoder based multi-task learning using probabilistic transcription. In *Proc. Interspeech*, pages 2073–2077, 2017.
- [80] 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.
- [81] 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.
- [82] Mark Hasegawa-Johnson, Preethi Jyothi, Wenda Chen, and Van Hai Do. Mismatched crowdsourcing: Mining latent skills to acquire speech transcriptions. In *Proceedings of Asilomar*, 2017.
- [83] 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.
- [84] Preethi Jyothi and Mark Hasegawa-Johnson. Low-resource grapheme-to-phoneme conversion using recurrent neural networks. In *Proc. ICASSP*, pages 5030–5034, 2017.
- [85] 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.
- [86] Mary Pietrowicz, Mark Hasegawa-Johnson, and Karrie Karahalios. Discovering dimensions of perceived vocal expression in semi-structured, unscripted oral history accoqunts. In *Proc. ICASSP*, pages 2901:1–4, 2017.
- [87] 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.
- [88] 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.

- [89] 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.
- [90] 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.
- [91] Daniel Morrow, Mark Hasegawa-Johnson, Thomas Huang, William Schuh, Rocio Garcia-Retamero, Renato Azevedo, Kuangxiao Gu, Yang Zhang, and Bidisha Roy. Multimedia formats can improve older adult comprehension of clinical test results: Implications for designing patient portals. In *28th APS Annual Convention (Association for Psychological Science)*, May 2016.
- [92] 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.
- [93] 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.
- [94] Amit Das and Mark Hasegawa-Johnson. An investigation on training deep neural networks using probabilistic transcription. In *Interspeech*, pages 3858–3862, 2016.
- [95] Amit Das, Preethi Jyothi, and Mark Hasegawa-Johnson. Automatic speech recognition using probabilistic transcriptions in swahili, amharic and dinka. In *Interspeech*, pages 3524–3528, 2016.
- [96] 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 *Interspeech*, pages 3863–3867, 2016.
- [97] 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.
- [98] 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.
- [99] 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.
- [100] 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.
- [101] Lav Varshney, Preethi Jyothi, and Mark Hasegawa-Johnson. Language coverage for mismatched crowdsourcing. In *Workshop on Information Theory and Applications*, 2016.
- [102] 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.
- [103] 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.
- [104] 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.
- [105] Raymond Yeh, Mark Hasegawa-Johnson, and Minh Do. Stable and symmetric filter convolutional neural network. In *Proc. ICASSP*, pages 2652–2656, 2016.

- [106] 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.
- [107] Renato F. L. Azevedo, Daniel Morrow, Mark Hasegawa-Johnson, Kuangxiao Gu, Dan Sobral, Thomas Huang, William Schuh, and Rocio Garcia-Retamero. Improving patient comprehension of numeric health information. In *Human Factors Conference*, 2015.
- [108] Amit Das and Mark Hasegawa-Johnson. Cross-lingual transfer learning during supervised training in low resource scenarios. In *Interspeech*, pages 3531–3535, 2015.
- [109] Lee Estelle, Lim Zhi Yi Vanessa, Ang Hui Shan, and Lim Boon Pang. Singapore hokkien speech recognition and applications. In *A*STAR research symposium*, 2015.
- [110] Preethi Jyothi and Mark Hasegawa-Johnson. Acquiring speech transcriptions using mismatched crowd-sourcing. In *Proc. AAAI*, pages 1263–1269, 2015.
- [111] 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 *Interspeech*, pages 3164–3168, 2015.
- [112] Preethi Jyothi and Mark Hasegawa-Johnson. Transcribing continuous speech using mismatched crowd-sourcing. In *Proc. Interspeech 2015*, pages 2774–2778, 2015.
- [113] Preethi Jyothi and Mark Hasegawa-Johnson. Transcribing continuous speech using mismatched crowd-sourcing. In *Interspeech*, pages 2774–2778, 2015.
- [114] Mary Pietrowicz, Mark Hasegawa-Johnson, and Karrie Karahalios. Acoustic correlates for perceived effort levels in expressive speech. In *Interspeech*, pages 3720–3724, 2015.
- [115] 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.
- [116] Jia Chen Ren, Mark Hasegawa-Johnson, and Lawrence Angrave. Classtranscribe. In *ICER Conference*, 2015.
- [117] Yang Zhang, Nasser Nasrabadi, and Mark Hasegawa-Johnson. Multichannel transient acoustic signal classification using task-driven dictionary with joint sparsity and beamforming. In *Proc. ICASSP*, pages 2591:1–5, 2015.
- [118] Yang Zhang, Zhijian Ou, and Mark Hasegawa-Johnson. Incorporating am-fm effect in voiced speech for probabilistic acoustic tube model. In *Proc. WASPAA*, 2015.
- [119] Sujeeth Bharadwaj and Mark Hasegawa-Johnson. A PAC-Bayesian approach to minimum perplexity language modeling. In *Proceedings of CoLing*, pages 130–140, 2014.
- [120] Xiayu Chen, Yang Zhang, and Mark Hasegawa-Jonson. An iterative approach to decision tree training for context dependent speech synthesis. In *Interspeech*, pages 2327–2331, 2014.
- [121] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. Automatic long audio alignment and confidence scoring for conversational arabic speech. In *The 9th edition of the Language Resources and Evaluation Conference (LREC 2014)*, Reykjavik, Iceland, 2014.
- [122] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. Development of a tv broadcasts speech recognition system for Qatari Arabic. In *The 9th edition of the Language Resources and Evaluation Conference (LREC 2014)*, pages 3057–3061, Reykjavik, Iceland, 2014.
- [123] 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.

- [124] Po-Sen Huang, Minje Kim, Paris Smaragdis, and Mark Hasegawa-Johnson. Deep learning for monaural speech separation. In *ICASSP*, 2014.
- [125] 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.
- [126] Preethi Jyothi, Jennifer Cole, Mark Hasegawa-Johnson, and Vandana Puri. An investigation of prosody in hindi narrative speech. In *Proceedings of Speech Prosody*, 2014.
- [127] 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.
- [128] 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.
- [129] Zhaowen Wang, Zhangyang Wang, Mark Moll, Po-Sen Huang, Devin Grady, Nasser Nasrabadi, Thomas Huang, Lydia Kavraki, and Mark Hasegawa-Johnson. Active planning, sensing and recognition using a resource-constrained discriminant pomdp. In *CVPR Multi-Sensor Fusion Workshop*, pages 740–747, 2014.
- [130] Yang Zhang, Zhijian Ou, and Mark Hasegawa-Johnson. Improvement of probabilistic acoustic tube model for speech decomposition. In *ICASSP*, 2014.
- [131] 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.
- [132] 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.
- [133] 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.
- [134] Sujeeth Bharadwaj, Mark Hasegawa-Johnson, Jitendra Ajmera, Om Deshmukh, and Ashish Verma. Sparse hidden markov models for purer clusters. In *Proc. ICASSP*, 2013.
- [135] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. Automatic long audio alignment for conversational arabic speech. In *Qatar Foundation Annual Research Conference*, 2013.
- [136] 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.
- [137] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. A framework for conversational arabic speech long audio alignment. In *Proc. 6th Language and Technology Conference (LTC 2013)*, pages 290–293, 2013.
- [138] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. A transfer learning approach for under-resourced arabic dialects speech recognition. In *Workshop on Less Resourced Languages, new technologies, new challenges and opportunities (LTC 2013)*, pages 60–64, 2013.
- [139] 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.
- [140] 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.

- [141] Sarah King and Mark Hasegawa-Johnson. Accurate speech segmentation by mimicking human auditory processing. In *Proc. ICASSP*, 2013.
- [142] Sujeeth Bharadwaj, Raman Arora, Karen Livescu, and Mark Hasegawa-Johnson. Multi-view acoustic feature learning using articulatory measurements. In *IWSML (Internat. Worksh. on Statistical Machine Learning for Sign. Process.)*, 2012.
- [143] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. A baseline speech recognition system for levantine colloquial arabic. In *Proceedings of ESOLEC*, 2012.
- [144] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. Hybrid pronunciation modeling for arabic large vocabulary speech recognition. In *Qatar Foundation Annual Research Forum*, 2012.
- [145] Camille Goudeseune. Effective browsing of long audio recordings. In *ACM International Workshop on Interactive Multimedia on Mobile and Portable Devices*, 2012.
- [146] 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.
- [147] 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.
- [148] Po-Sen Huang, Scott Deeann Chen, Paris Smaragdis, and Mark Hasegawa-Johnson. Singing-voice separation from monaural recordings using robust principal component analysis. In *ICASSP*, 2012.
- [149] Po-Sen Huang and Mark Hasegawa-Johnson. Cross-dialectal data transferring for gaussian mixture model training in arabic speech recognition. In *International Conference on Arabic Language Processing CITALA*, pages 119–122, 2012.
- [150] 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.
- [151] Po-Sen Huang, Robert Mertens, Ajay Divakaran, Gerald Friedland, and Mark Hasegawa-Johnson. How to put it into words—using random forests to extract symbol level descriptions from audio content for concept detection. In *ICASSP*, 2012.
- [152] Po-Sen Huang, Jianchao Yang, Mark Hasegawa-Johnson, Feng Liang, and Thomas S. Huang. Pooling robust shift-invariant sparse representations of acoustic signals. In *Interspeech*, pages 2518–2521, 2012.
- [153] Sarah King and Mark Hasegawa-Johnson. Detection of acoustic-phonetic landmarks in mismatched conditions using a biomimetic model of human auditory processing. In *CoLing*, pages 589–598, 2012.
- [154] Kai-Hsiang Lin, Xiaodan Zhuang, Camille Goudeseune, Sarah King, Mark Hasegawa-Johnson, and Thomas Huang. Improving faster-than-real-time human acoustic event detection by saliency-maximized audio visualization. In *ICASSP*, pages 2277–2280, 2012.
- [155] Tim Mahrt, Jennifer Cole, Margaret Fleck, and Mark Hasegawa-Johnson. Accounting for speaker variation in the production of prominence using the bayesian information criterion. In *Speech Prosody*, 2012.
- [156] Tim Mahrt, Jennifer Cole, Margaret Fleck, and Mark Hasegawa-Johnson. F0 and the perception of prominence. In *Proc. Interspeech 2012*, pages 2422–2425, 2012.
- [157] Ali Sakr and Mark Hasegawa-Johnson. Topic modeling of phonetic latin-spelled arabic for the relative analysis of genre-dependent and dialect-dependent variation. In *CITALA*, pages 153–158, 2012.
- [158] 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.

- [159] Arthur Kantor and Mark Hasegawa-Johnson. Hmm-based pronunciation dictionary generation. In *Workshop on New Tools and Methods for Very Large Scale Phonetics Research*, University of Pennsylvania, Jan. 2011.
- [160] 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.
- [161] 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.
- [162] Mohamed Elmahdy, Mark Hasegawa-Johnson, Eiman Mustafawi, Rehab Duwairi, and Wolfgang Minker. Challenges and techniques for dialectal arabic speech recognition and machine translation. In *Qatar Foundation Annual Research Forum*, page 244, 2011.
- [163] Mark Hasegawa-Johnson, Camille Goudeseune, Jennifer Cole, Hank Kaczmarski, Heejin Kim, Sarah King, Timothy Mahrt, Jui-Ting Huang, Xiaodan Zhuang, Kai-Hsiang Lin, Harsh Vardhan Sharma, Zhen Li, and Thomas S. Huang. Multimodal speech and audio user interfaces for K-12 outreach. In *APSIPA*, pages 256:1–8, 2011.
- [164] Po-Sen Huang, Thyagaraju Damarla, and Mark Hasegawa-Johnson. Multi-sensory features for personnel detection at border crossings. In *Fusion*, 2011.
- [165] Po-Sen Huang, Mark Hasegawa-Johnson, and Thyagaraju Damarla. Exemplar selection methods to distinguish human from animal footsteps. In *Second Annual Human and Light Vehicle Detection Workshop*, pages 14:1–10, Maryland, 2011.
- [166] Po-Sen Huang, Xiaodan Zhuang, and Mark Hasegawa-Johnson. Improving acoustic event detection using generalizable visual features and multi-modality modeling. In *ICASSP*, pages 349–352, 2011.
- [167] 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.
- [168] 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.
- [169] 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.
- [170] 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.
- [171] David Harwath and Mark Hasegawa-Johnson. Phonetic landmark detection for automatic language identification. In *Speech Prosody*, pages 100231:1–4, 2010.
- [172] Chi Hu, Xiaodan Zhuang, and Mark Hasegawa-Johnson. Fsm-based pronunciation modeling using articulatory phonological code. In *Proceedings of Interspeech*, pages 2274–2277, 2010.
- [173] Jui-Ting Huang and Mark Hasegawa-Johnson. Semi-supervised training of gaussian mixture models by conditional entropy minimization. In *Proceedings of Interspeech*, pages 1353–1356, 2010.
- [174] Jui-Ting Huang, Po-Sen Huang, Yoonsook Mo, Mark Hasegawa-Johnson, and Jennifer Cole. Prosody-dependent acoustic modeling using variable-parameter hidden markov models. In *Speech Prosody*, pages 100623:1–4, 2010.

- [175] Heejin Kim, Mark Hasegawa-Johnson, and Adrienne Perlman. Acoustic cues to lexical stress in spastic dysarthria. In *Speech Prosody*, pages 100891:1–4, 2010.
- [176] Heejin Kim, Panyang Rong, Torrey M. Loucks, and Mark Hasegawa-Johnson. Kinematic analysis of tongue movement control in spastic dysarthria. In *Proceedings of Interspeech*, pages 2578–2581, 2010.
- [177] Lae-Hoon Kim and Mark Hasegawa-Johnson. Toward overcoming fundamental limitation in frequency-domain blind source separation for reverberant speech mixtures. In *Proceedings of Asilomar*, 2010.
- [178] Lae-Hoon Kim, Mark Hasegawa-Johnson, Gerasimos Potamianos, and Vit Libal. Joint estimation of doa and speech based on em beamforming. In *ICASSP*, 2010.
- [179] Lae-Hoon Kim, Kyung-Tae Kim, and Mark Hasegawa-Johnson. Robust automatic speech recognition with decoder oriented ideal binary mask estimation. In *Proceedings of Interspeech*, pages 2066–2069, 2010.
- [180] Yoonsook Mo, Jennifer Cole, and Mark Hasegawa-Johnson. Prosodic effects on temporal structure of monosyllabic cvc words in american english. In *Speech Prosody*, pages 100208:1–4, 2010.
- [181] Hosung Nam, Vikramjit Mitra, Mark Tiede, Elliot Saltzman, Louis Goldstein, Carol Espy-Wilson, and Mark Hasegawa-Johnson. A procedure for estimating gestural scores from natural speech. In *Proceedings of Interspeech*, pages 30–33, 2010.
- [182] Harsh Vardhan Sharma and Mark Hasegawa-Johnson. State transition interpolation and map adaptation for hmm-based dysarthric speech recognition. In *HLT/NAACL Workshop on Speech and Language Processing for Assistive Technology (SLPAT)*, pages 72–79, 2010.
- [183] Hao Tang, Mark Hasegawa-Johnson, and Thomas S. Huang. Non-frontal view facial expression recognition. In *ICME*, pages 1202–1207, 2010.
- [184] Hao Tang, Mark Hasegawa-Johnson, and Thomas S. Huang. Toward robust learning of the gaussian mixture state emission densities for hidden markov models. In *ICASSP*, 2010.
- [185] Su-Youn Yoon, Mark Hasegawa-Johnson, and Richard Sproat. Landmark-based automated pronunciation error detection. In *Proceedings of Interspeech*, pages 614–617, 2010.
- [186] Xiaodan Zhuang, Lijuan Wang, Frank Soong, and Mark Hasegawa-Johnson. A minimum converted trajectory error (MCTE) approach to high quality speech-to-lips conversion. In *Proceedings of Interspeech*, pages 1736–1739, 2010.
- [187] Yoonsook Mo, Jennifer Cole, and Mark Hasegawa-Johnson. Prosodic effects on vowel production: evidence from formant structure. In *Interspeech*, pages 2535–2538, Brighton, September 2009.
- [188] I. Yücel Özbek, Mark Hasegawa-Johnson, and Mübeccel Demirekler. Formant trajectories for acoustic-to-articulatory inversion. In *Interspeech*, pages 2807–2810, Brighton, September 2009.
- [189] Harsh Vardhan Sharma, Mark Hasegawa-Johnson, Jon Gunderson, and Adrienne Perlman. Universal access: Speech recognition for talkers with spastic dysarthria. In *Interspeech*, pages 1451–1454, Brighton, September 2009.
- [190] Su-Youn Yoon, Mark Hasegawa-Johnson, and Richard Sproat. Automated pronunciation scoring using confidence scoring and landmark-based svm. In *Interspeech*, pages 1903–1906, Brighton, September 2009.
- [191] Xiaodan Zhuang, Hosung Nam, Mark Hasegawa-Johnson, Louis Goldstein, and Elliot Saltzman. Articulatory phonological code for word recognition. In *Interspeech*, pages 2763–2766, Brighton, September 2009.
- [192] Lae-Hoon Kim and Mark Hasegawa-Johnson. Optimal multi-microphone speech enhancement in cars. In *DSP in Cars workshop*, Dallas, July 2009.

- [193] 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.
- [194] Mark Hasegawa-Johnson, Camille Goudeseune, Kai-Hsiang Lin, David Cohen, Xi Zhou, Xiaodan Zhuang, Kyungtae Kim, Hank Kaczmarski, and Thomas Huang. Visual analytics for audio. In *NIPS Workshop on Visual Analytics*, 2009.
- [195] Jui-Ting Huang and Mark Hasegawa-Johnson. On semi-supervised learning of gaussian mixture models for phonetic classification. In *NAACL HLT Workshop on Semi-Supervised Learning*, pages 75–83, 2009.
- [196] Jui-Ting Huang, Xi Zhou, Mark Hasegawa-Johnson, and Thomas Huang. Kernel metric learning for phonetic classification. In *ASRU*, pages 141–145, 2009.
- [197] Hao Tang, Stephen M. Chu, Mark Hasegawa-Johnson, and Thomas S. Huang. Emotion recognition from speech via boosted gaussian mixture models. In *International Conference on Multimedia & Expo (ICME’09)*, pages 294–297, 2009.
- [198] Su-Youn Yoon, Mark Hasegawa-Johnson, and Richard Sproat. Automated pronunciation scoring for l2 english learners. In *CALICO workshop*, 2009.
- [199] 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.
- [200] Xiaodan Zhuang, Jui-Ting Huang, and Mark Hasegawa-Johnson. Speech retrieval in unknown languages: a pilot study. In *NAACL HLT Cross-Lingual Information Access Workshop (CLIAWS)*, pages 3–11, 2009.
- [201] Xiaodan Zhuang, Xi Zhou, Mark A. Hasegawa-Johnson, and Thomas S. Huang. Efficient object localization with gaussianized vector representation. In *IMCE*, pages 89–96, 2009.
- [202] Jui-Ting Huang and Mark Hasegawa-Johnson. Maximum mutual information estimation with unlabeled data for phonetic classification. In *Proc. Interspeech*, 2008.
- [203] Jui-Ting Huang and Mark Hasegawa-Johnson. Unsupervised prosodic break detection in mandarin speech. In *SpeechProsody*, pages 165–168, 2008.
- [204] Arthur Kantor and Mark Hasegawa-Johnson. Stream weight tuning in dynamic bayesian networks. In *Proc. ICASSP*, pages 4525–4528, 2008.
- [205] Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas Huang, Kenneth Watkin, and Simone Frame. Dysarthric speech database for universal access research. In *Interspeech*, pages 1741–1744, 2008.
- [206] 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.
- [207] Bryce Lobdell, Mark Hasegawa-Johnson, and Jont B. Allen. Human speech perception and feature extraction. In *Interspeech*, pages 1797–1800, 2008.
- [208] 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.
- [209] Hao Tang, Yun Fu, Jilin Tu, Thomas Huang, and Mark Hasegawa-Johnson. Eava: A 3d emotive audio-visual avatar. In *IEEE Workshop on Applications of Computer Vision (IEEE WACV)*, pages 1–6, 2008.
- [210] 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.

- [211] Hao Tang, Xi Zhou, Matthias Odisio, Mark Hasegawa-Johnson, and Thomas Huang. Two-stage prosody prediction for emotional text-to-speech synthesis. In *Interspeech*, pages 2138–2141, 2008.
- [212] 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.
- [213] Taejin Yoon, Jennifer Cole, and Mark Hasegawa-Johnson. Detecting non-modal phonation in telephone speech. In *SpeechProsody*, pages 33–36, 2008.
- [214] Xi Zhou, Xiaodan Zhuang, Hao Tang, Mark Hasegawa-Johnson, and Thomas Huang. A novel gaussianized vector representation for natural scene categorization. In *ICPR*, pages 1–4, 2008.
- [215] Xi Zhou, Xiaodan Zhuang, Shuicheng Yan, Shih-Fu Chang, Mark Hasegawa-Johnson, and Thomas S. Huang. Sift-bag kernel for video event analysis. In *ACM Multimedia*, pages 229–238, 2008.
- [216] Xiaodan Zhuang and Mark Hasegawa-Johnson. Towards interpretation of creakiness in switchboard. In *SpeechProsody*, pages 37–40, 2008.
- [217] Xiaodan Zhuang, Hosung Nam, Mark Hasegawa-Johnson, Louis Goldstein, and Elliot Saltzman. The entropy of articulatory phonological code: Recognizing gestures from tract variables. In *Interspeech*, pages 1489–1492, 2008.
- [218] Xiaodan Zhuang, Hosung Nam, Mark Hasegawa-Johnson, Louis Goldstein, and Elliot Saltzman. The entropy of articulatory phonological code: Recognizing gestures from tract variables. In *Interspeech*, pages 1489–1492, 2008.
- [219] Xiaodan Zhuang, Xi Zhou, Mark Hasegawa-Johnson, and Thomas Huang. Face age estimation using patch-based hidden markov model supervectors. In *ICPR*, pages 1–4, 2008.
- [220] Xiaodan Zhuang, Xi Zhou, Thomas S. Huang, and Mark Hasegawa-Johnson. Feature analysis and selection for acoustic event detection. In *ICASSP*, pages 17–20, 2008.
- [221] Mark Hasegawa-Johnson, Karen Livescu, Partha Lal, and Kate Saenko. Audiovisual speech recognition with articulator positions as hidden variables. In *Proc. International Congress on Phonetic Sciences (ICPhS)*, pages 1719:297–302, Saarbrücken, August 2007.
- [222] 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.
- [223] Taejin Yoon, Jennifer Cole, and Mark Hasegawa-Johnson. On the edge: Acoustic cues to layered prosodic domains. In *Proc. International Congress on Phonetic Sciences (ICPhS)*, pages 1264:1017–1020, Saarbrücken, August 2007.
- [224] Bowon Lee and Mark Hasegawa-Johnson. Minimum mean squared error a posteriori estimation of high variance vehicular noise. In *2007 Biennial on DSP for In-Vehicle and Mobile Systems*, Istanbul, June 2007.
- [225] 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.
- [226] Taejin Yoon, Jennifer Cole, and Mark Hasegawa-Johnson. On the edge: Acoustic cues to layered prosodic domains. In *81st Annual Meeting of the Linguistic Society of America*, Anaheim, CA, January 2007.

- [227] Yun Fu, Xi Zhou, Ming Liu, Mark Hasegawa-Johnson, and Thomas S. Huang. Lipreading by locality discriminant graph. In *IEEE International Conference on Image Processing (ICIP)*, pages III:325–8, 2007.
- [228] Mark Hasegawa-Johnson. Multi-stream approach to audiovisual automatic speech recognition. In *IEEE 9th Workshop on Multimedia Signal Processing (MMSP)*, pages 328–331, 2007.
- [229] Ming Liu, Zhengyou Zhang, Mark Hasegawa-Johnson, and Thomas Huang. Exploring discriminative learning for text-independent speaker recognition. In *ICME*, pages 56–59, 2007.
- [230] Xi Zhou, Yu Fun, Ming Liu, Mark Hasegawa-Johnson, and Thomas Huang. Robust analysis and weighting on mfcc components for speech recognition and speaker identification. In *International Conference on Multimedia and Expo*, pages 188–191, 2007.
- [231] Weimo Zhu, Mark Hasegawa-Johnson, Karen Chapman-Novakofski, and Arthur Kantor. Cellphone-based nutrition e-diary. In *National Nutrient Database Conference*, 2007.
- [232] 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.
- [233] Rahul Chitturi and Mark Hasegawa-Johnson. Novel time-domain multi-class svms for landmark detection. In *Interspeech*, pages paper 1904–Thu1CaP.14, September 2006.
- [234] Mark Hasegawa-Johnson. Dealing with acoustic noise. part iii: Video. In *tutorial presentation given at WS06, Center for Language and Speech Processing*, July 2006.
- [235] Mark Hasegawa-Johnson, Jonathan Gunderson, Adrienne Perlman, and Thomas Huang. Hmm-based and svm-based recognition of the speech of talkers with spastic dysarthria. In *ICASSP*, pages III:1060–3, May 2006.
- [236] 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.
- [237] Soo-Eun Chang, Kirk I. Erickson, Nicoline G. Ambrose, Mark Hasegawa-Johnson, and C.L. Ludlow. Deficient white matter development in left hemisphere speech-language regions in children who stutter. In *Society for Neuroscience*, Atlanta, GA, 2006.
- [238] 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.
- [239] Heejin Kim, Taejin Yoon, Jennifer Cole, and Mark Hasegawa-Johnson. Acoustic differentiation of l- and l-1% in switchboard and radio news speech. In *Proceedings of Speech Prosody*, Dresden, 2006.
- [240] Rajiv Reddy and Mark Hasegawa-Johnson. Analysis of pitch contours in repetition-disfluency using stem-ml. In *Midwest Computational Linguistics Colloquium*, 2006.
- [241] Sarah Borys and Mark Hasegawa-Johnson. Distinctive feature based SVM discriminant features for improvements to phone recognition on telephone band speech. In *ISCA Interspeech*, October 2005.
- [242] 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.
- [243] Mark Hasegawa-Johnson, James Baker, Sarah Borys, Ken Chen, Emily Coogan, Steven Greenberg, Amit Juneja, Katrin Kirchhoff, Karen Livescu, Srividya Mohan, Jennifer Muller, Kemal Sönmez, and Tianyu Wang. Landmark-based speech recognition: Report of the 2004 johns hopkins summer workshop. In *ICASSP*, pages 1213–1216, March 2005.

- [244] 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.
- [245] 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.
- [246] Arthur Kantor, Weimo Zhu, and Mark Hasegawa-Johnson. Restricted domain speech classification using automatic transcription and svms. In *Midwest Computational Linguistics Colloquium*, 2005.
- [247] Yeojin Kim and Mark Hasegawa-Johnson. Phonetic segment rescoring using svms. In *Midwest Computational Linguistics Colloquium*, Columbus, OH, 2005.
- [248] Tong Zhang, Mark Hasegawa-Johnson, and Stephen E. Levinson. A hybrid model for spontaneous speech understanding. In *AAAI*, pages 10.1.1.80.879:1–8, 2005.
- [249] Soo-Eun Chang, Nicoline Ambrose, and Mark Hasegawa-Johnson. An mri (dti) study on children with persistent developmental stuttering. In *ASHA Convention*, November 2004.
- [250] Sarah Borys, Mark Hasegawa-Johnson, Ken Chen, and Aaron Cohen. Modeling and recognition of phonetic and prosodic factors for improvements to acoustic speech recognition models. In *Interspeech*, pages 3013–3016, October 2004.
- [251] Sarah Borys, Mark Hasegawa-Johnson, Ken Chen, and Aaron Cohen. Modeling and recognition of phonetic and prosodic factors for improvements to acoustic speech recognition models. In *Interspeech*, pages 3013–3016, October 2004.
- [252] Ken Chen and Mark Hasegawa-Johnson. Modeling pronunciation variation using artificial neural networks for english spontaneous speech. In *Interspeech*, pages 400–403, October 2004.
- [253] Mital Gandhi and Mark Hasegawa-Johnson. Source separation using particle filters. In *Interspeech*, pages 2673–2676, October 2004.
- [254] Mark Hasegawa-Johnson and Ameya Deoras. A factorial hmm approach to robust isolated digit recognition in background music. In *Interspeech*, pages 2093–2096, October 2004.
- [255] Mark Hasegawa-Johnson, Stephen E. Levinson, and Tong Zhang. Children’s emotion recognition in an intelligent tutoring scenario. In *Interspeech*, pages 1441–1444, October 2004.
- [256] Bowon Lee, Mark Hasegawa-Johnson, Camille Goudeseune, Suketu Kamdar, Sarah Borys, Ming Liu, and Thomas Huang. Avicar: Audio-visual speech corpus in a car environment. In *Interspeech*, pages 380–383, October 2004.
- [257] Tae-Jin Yoon, Sandra Chavarria, Jennifer Cole, and Mark Hasegawa-Johnson. Intertranscriber reliability of prosodic labeling on telephone conversation using tobi. In *Interspeech*, pages 2729–2732, October 2004.
- [258] Yanli Zheng, Mark Hasegawa-Johnson, and Sarah Borys. Stop consonant classification by dynamic formant trajectory. In *Interspeech*, pages 396–399, October 2004.
- [259] Mark Hasegawa-Johnson, Sarah Borys, and Ken Chen. Experiments in landmark-based speech recognition. In *Sound to Sense: Workshop in Honor of Kenneth N. Stevens*, June 2004.
- [260] Mark Hasegawa-Johnson, Jennifer Cole, Chilin Shih, Ken Chen, Aaron Cohen, Sandra Chavarria, Heejin Kim, Taejin Yoon, Sarah Borys, and Jeung-Yoon Choi. Speech recognition models of the interdependence among syntax, prosody, and segmental acoustics. In *HLT/NAACL Workshop on Higher-Level Knowledge in Automatic Speech Recognition and Understanding*, pages 56–63, May 2004.

- [261] 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.
- [262] Ken Chen and Mark Hasegawa-Johnson. How prosody improves word recognition. In *Speech Prosody*, pages 583–586, Nara, Japan, March 2004.
- [263] Ken Chen, Mark Hasegawa-Johnson, Aaron Cohen, and Jennifer Cole. A maximum likelihood prosody recognizer. In *SpeechProsody*, pages 509–512, Nara, Japan, March 2004.
- [264] 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.
- [265] Yuexi Ren, Sung-Suk Kim, Mark Hasegawa-Johnson, and Jennifer Cole. Speaker-independent automatic detection of pitch accent. In *Proc. Speech Prosody 2004*, pages 521–524, Nara, Japan, March 2004.
- [266] Ken Chen and Mark Hasegawa-Johnson. An automatic prosody labeling system using ann-based syntactic-prosodic model and gmm-based acoustic-prosodic model. In *ICASSP*, 2004.
- [267] Ameya Deoras and Mark Hasegawa-Johnson. A factorial hmm approach to simultaneous recognition of isolated digits spoken by multiple talkers on one audio channel. In *ICASSP*, 2004.
- [268] Mark Hasegawa-Johnson, Stephen Levinson, and Tong Zhang. Automatic detection of contrast for speech understanding. In *Proc. Interspeech 2004*, pages 581–584, 2004.
- [269] Mark Hasegawa-Johnson, Stephen Levinson, and Tong Zhang. Automatic detection of contrast for speech understanding. In *Proc. Interspeech 2004*, pages 581–584, 2004.
- [270] Yuexi Ren, Mark Hasegawa-Johnson, and Stephen E. Levinson. Semantic analysis for a speech user interface in an intelligent-tutoring system. In *Intl. Conf. on Intelligent User Interfaces*, Madeira, Portugal, 2004.
- [271] 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.
- [272] Yanli Zheng and Mark Hasegawa-Johnson. Formant tracking by mixture state particle filter. In *ICASSP*, 2004.
- [273] 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 *Interspeech*, pages 393–396, September 2003.
- [274] Mohammed Kamal Omar and Mark Hasegawa-Johnson. Maximum conditional mutual information projection for speech recognition. In *Interspeech*, pages 505–508, September 2003.
- [275] Mohammed Kamal Omar and Mark Hasegawa-Johnson. Non-linear maximum likelihood feature transformation for speech recognition. In *Interspeech*, pages 2497–2500, September 2003.
- [276] Yanli Zheng and Mark Hasegawa-Johnson. Particle filtering approach to bayesian formant tracking. In *IEEE Workshop on Statistical Signal Processing*, pages 581–584, September 2003.
- [277] Tong Zhang, Mark Hasegawa-Johnson, and Stephen E. Levinson. Mental state detection of dialogue system users via spoken language. In *ISCA/IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR)*, pages MAP17.1–4, April 2003.
- [278] Yanli Zheng and Mark Hasegawa-Johnson. Acoustic segmentation using switching state kalman filter. In *ICASSP*, pages I:752–755, April 2003.

- [279] Sarah Borys, Mark Hasegawa-Johnson, and Jennifer Cole. The importance of prosodic factors in phoneme modeling with applications to speech recognition. In *ACL Student Session*, 2003.
- [280] Ken Chen and Mark Hasegawa-Johnson. Improving the robustness of prosody dependent language modeling based on prosody syntax cross-correlation. In *ASRU*, 2003.
- [281] 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.
- [282] 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.
- [283] 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.
- [284] Mark Hasegawa-Johnson. Bayesian learning for models of human speech perception. In *IEEE Workshop on Statistical Signal Processing*, pages 393–396, St. Louis, MO, 2003.
- [285] 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.
- [286] Mohammed Kamal Omar and Mark Hasegawa-Johnson. Non-linear independent component analysis for speech recognition. In *International Conference on Computer, Communication and Control Technologies (CCCT '03)*, 2003.
- [287] Mohammed Kamal Omar and Mark Hasegawa-Johnson. Strong-sense class-dependent features for statistical recognition. In *IEEE Workshop on Statistical Signal Processing*, pages 473–476, St. Louis, MO, 2003.
- [288] Tong Zhang, Mark Hasegawa-Johnson, and Stephen E. Levinson. An empathic-tutoring system using spoken language. In *Australian conference on computer-human interaction (OZCHI)*, pages 498–501, 2003.
- [289] 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 *Interspeech*, pages 2129–2132, Denver, CO, September 2002.
- [290] Stephen E. Levinson, Thomas S. Huang, Mark A. Hasegawa-Johnson, Ken Chen, Stephen Chu, Ashutosh Garg, Zhinian Jing, Danfeng Li, J. Lin, Mohammed Kamal Omar, and Z. Wen. Multi-modal dialog systems research at illinois. In *ARPA Workshop on Multimodal Speech Recognition and SPINE*, June 2002.
- [291] Zhinian Jing and Mark Hasegawa-Johnson. Auditory-modeling inspired methods of feature extraction for robust automatic speech recognition. In *ICASSP*, page IV:4176, May 2002.
- [292] 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.
- [293] 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.
- [294] Mohammed K. Omar, Mark Hasegawa-Johnson, and Stephen E. Levinson. Gaussian mixture models of phonetic boundaries for speech recognition. In *ASRU*, 2001.
- [295] Mark Hasegawa-Johnson. Multivariate-state hidden markov models for simultaneous transcription of phones and formants. In *ICASSP*, pages 1323–1326, Istanbul, 2000.

- [296] 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.
- [297] 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.
- [298] Mark Hasegawa-Johnson. Combining magnetic resonance image planes in the fourier domain for improved spatial resolution. In *International Conference On Signal Processing Applications and Technology*, pages 81.1–5, Orlando, FL, 1999.
- [299] Mark Hasegawa-Johnson, Jul Cha, and Katherine Haker. Ctmredit: A matlab-based tool for segmenting and interpolating mri and ct images in three orthogonal planes. In *21st Annual International Conference of the IEEE/EMBS Society*, page 1170, 1999.
- [300] Mark Hasegawa-Johnson, Jul Cha, Shamala Pizza, and Katherine Haker. Ctmredit: A case study in human-computer interface design. In *International Conference On Public Participation and Information Technology*, pages 575–584, Lisbon, 1999.
- [301] Mark Johnson. Automatic context-sensitive measurement of the acoustic correlates of distinctive features. In *ICSLP*, pages 1639–1643, Yokohama, 1994.
- [302] 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.
- [303] 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.
- [304] 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.
- [305] 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.
- [306] Mark A. Johnson and Tomohiko Taniguchi. Pitch-orthogonal code-excited LPC. In *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 542–546, San Diego, CA, 1990.
- [307] 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.

Journal Articles and JASA Abstracts

- [1] 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.
- [2] Heting Gao, Xiaoxuan Wang, Sunghun Kang, Rusty Mina, Dias Issa, John Harvill, Leda Sari, Mark Hasegawa-Johnson, and Chang D. Yoo. Seamless equal accuracy ratio for inclusive ctc speech recognition. *Speech Communication*, 136:76–83, 2022.
- [3] 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.

- [4] 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.
- [5] 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.
- [6] Leda Sari, Mark Hasegawa-Johnson, and Chang D. Yoo. Counterfactually fair automatic speech recognition. *IEEE Transactions on Audio, Speech, and Language*, 29:3515–3525, 2021.
- [7] Odette Scharenborg and Mark Hasegawa-Johnson. Position paper: Brain signal-based dialogue systems. *Lecture Notes in Computer Science*, 714, 3 2021.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] Xiang Kong, Preethi Jyothi, and Mark Hasegawa-Johnson. Performance improvement of probabilistic transcriptions with language-specific constraints. *Procedia Computer Science*, 81:30–36, 2016.

- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] Austin Chen and Mark Hasegawa-Johnson. Mixed stereo audio classification using a stereo-input mixed-to-panned level feature. *IEEE Trans. Speech and Audio Processing*, 22(12):2025–2033, 2014.
- [24] 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.
- [25] 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.
- [26] Kyungtae Kim, Kai-Hsiang Lin, Dirk B Walther, Mark A Hasegawa-Johnson, and Thomas S Huang. Automatic detection of auditory salience with optimized linear filters derived from human annotation. *Pattern Recognition Letters*, 38(1):78–85, 2013.
- [27] Kai-Hsiang Lin, Xiaodan Zhuang, Camille Goudeseune, Sarah King, Mark Hasegawa-Johnson, and Thomas S. Huang. Saliency-maximized audio visualization and efficient audio-visual browsing for faster-than-real-time human acoustic event detection. *ACM Transactions on Applied Perception*, 2013.
- [28] 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.
- [29] Mohamed Elmahdy, Mark Hasegawa-Johnson, and Eiman Mustafawi. Hybrid phonemic and graphemic modeling for arabic speech recognition. *International Journal of Computational Linguistics*, 3(1):88–96, 2012.
- [30] Shobhit Mathur, Marshall Scott Poole, Feniosky Pena-Mora, Mark Hasegawa-Johnson, and Noshir Contractor. Detecting interaction links in a collaborating group using manually annotated data. *Social Networks*, 2012.
- [31] 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.
- [32] 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.
- [33] Mark Hasegawa-Johnson, Jui-Ting Huang, Sarah King, and Xi Zhou. Normalized recognition of speech and audio events. *Journal of the Acoustical Society of America*, 130:2524, 2011.
- [34] Mark Hasegawa-Johnson, Jui-Ting Huang, and Xiaodan Zhuang. Semi-supervised learning for speech and audio processing. *Journal of the Acoustical Society of America*, 130:2408, 2011.

- [35] 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.
- [36] Heejin Kim, Mark Hasegawa-Johnson, and Adrienne Perlman. Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, 63(4):187–194, 2011.
- [37] 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.
- [38] 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.
- [39] Xiaodan Zhuang, Xi Zhou, Mark A. Hasegawa-Johnson, and Thomas S. Huang. Real-world acoustic event detection. *Pattern Recognition Letters*, 31(2):1543–1551, Sep. 2010.
- [40] Xi Zhou, Xiaodan Zhuang, Hao Tang, Mark A. Hasegawa-Johnson, and Thomas S. Huang. Novel gaussianized vector representation for improved natural scene categorization. *Pattern Recognition Letters*, 31(8):702–708, Jun. 2010.
- [41] Heejin Kim, Katie Martin, Mark Hasegawa-Johnson, and Adrienne Perlman. Frequency of consonant articulation errors in dysarthric speech. *Clinical Linguistics & Phonetics*, 24(10):759–770, 2010.
- [42] Lae-Hoon Kim, Kyungtae Kim, and Mark Hasegawa-Johnson. Speech enhancement beyond minimum mean squared error with perceptual noise shaping. *JASA*, 127(3):1817, 2010.
- [43] Bryce E Lobdell, Jont B Allen, and Mark A Hasegawa-Johnson. Intelligibility predictors and neural representation of speech. *Speech Communication*, 2010.
- [44] 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.
- [45] Mark Hasegawa-Johnson. Tutorial: Pattern recognition in signal processing. *JASA*, 125:2698, 2009.
- [46] Mark Hasegawa-Johnson, Xiaodan Zhuang, Xi Zhou, Camille Goudeseune, and Thomas S. Huang. Adaptation of tandem hmms for non-speech audio event detection. *JASA*, 125:2730, 2009.
- [47] Yoonsook Mo, Jennifer Cole, and Mark Hasegawa-Johnson. How do ordinary listeners perceive prosodic prominence? syntagmatic vs. paradigmatic comparison. *JASA*, 125(4):2572, 2009.
- [48] Su-Youn Yoon, Lisa Pierce, Amanda Huensch, Eric Juul, Samantha Perkins, Richard Sproat, and Mark Hasegawa-Johnson. Construction of a rated speech corpus of 12 learners’ speech. *CALICO journal*, 2009.
- [49] Mark Hasegawa-Johnson, Jennifer Cole, Ken Chen, Partha Lal, Amit Juneja, Taejin Yoon, Sarah Borys, and Xiaodan Zhuang. Prosodically organized automatic speech recognition. *Language and Linguistics Monograph Series*, A25:101–128, 2008.
- [50] Lae-Hoon Kim, Mark Hasegawa-Johnson, Jun-Seok Lim, and Koeng-Mo Sung. Acoustic model for robustness analysis of optimal multipoint room equalization. *JASA*, 123(4):2043–2053, 2008.
- [51] 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.
- [52] Taejin Yoon, Xiaodan Zhuang, Jennifer Cole, and Mark Hasegawa-Johnson. Voice quality dependent speech recognition. *Language and Linguistics Monograph Series*, A25:77–100, 2008.
- [53] Xi Zhou, Xiaodan Zhuang, Ming Lui, Hao Tang, Mark Hasegawa-Johnson, and Thomas Huang. Hmm-based acoustic event detection with adaboost feature selection. *Lecture Notes in Computer Science*, 4625:345–353, 2008.

- [54] Soo-Eun Chang, Noline Ambrose, Kirk Erickson, and Mark Hasegawa-Johnson. Brain anatomy differences in childhood stuttering. *Neuroimage*, 2007.
- [55] Jennifer Cole, Heejin Kim, Hansook Choi, and Mark Hasegawa-Johnson. Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech. *J Phonetics*, 35:180–209, 2007.
- [56] Jennifer Cole, Yoonsook Mo, and Mark Hasegawa-Johnson. Signal-based and expectation-based factors in the perception of prosodic prominence. *Journal of Laboratory Phonology*, 2007.
- [57] Tong Zhang, Mark Hasegawa-Johnson, and Stephen E. Levinson. Extraction of pragmatic and semantic salience from spontaneous spoken english. *Speech Communication*, 2007.
- [58] Ken Chen, Mark Hasegawa-Johnson, Aaron Cohen, Sarah Borys, Sung-Suk Kim, Jennifer Cole, and Jeung-Yoon Choi. Prosody dependent speech recognition on radio news corpus of american english. *IEEE Transactions on Speech and Audio Processing*, 14(1):232–245, 2006.
- [59] Tong Zhang, Mark Hasegawa-Johnson, and Stephen E. Levinson. Cognitive state classification in a spoken tutorial dialogue system. *Speech Communication*, 48(6), 2006.
- [60] 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.
- [61] Mark Hasegawa-Johnson, Ken Chen, Jennifer Cole, Sarah Borys, Sung-Suk Kim, Aaron Cohen, Tong Zhang, Jeung-Yoon Choi, Heejin Kim, Taejin Yoon, and Sandra Chavarria. Simultaneous recognition of words and prosody in the boston university radio speech corpus. *Speech Communication*, 46(3-4):418–439, 2005.
- [62] Tae-Jin Yoon, Cole, Jennifer, Mark Hasegawa-Johnson, and Chilin Shih. Acoustic correlates of non-modal phonation in telephone speech. *The Journal of the Acoustical Society of America*, 117(4):2621, 2005.
- [63] Sung-Suk Kim, Mark Hasegawa-Johnson, and Ken Chen. Automatic recognition of pitch movements using multilayer perceptron and time-delay recursive neural network. *IEEE Signal Processing Letters*, 11(7):645–648, 2004.
- [64] Mohammad Kamal Omar and Mark Hasegawa-Johnson. Model enforcement: A unified feature transformation framework for classification and recognition. *IEEE Transactions on Signal Processing*, 52(10):2701–2710, 2004.
- [65] Yanli Zheng, Mark Hasegawa-Johnson, and Shamala Pizza. Analysis of the three-dimensional tongue shape using a three-index factor analysis model. *Journal of the Acoustical Society of America*, 113(1):478–486, January 2003.
- [66] Sarah Borys, Mark Hasegawa-Johnson, and Jennifer Cole. Prosody as a conditioning variable in speech recognition. *Illinois Journal of Undergraduate Research*, 2003.
- [67] Mark Hasegawa-Johnson, Shamala Pizza, Abeer Alwan, Jul Cha, and Katherine Haker. Vowel category dependence of the relationship between palate height, tongue height, and oral area. *Journal of Speech, Language, and Hearing Research*, 46(3):738–753, 2003.
- [68] Bowon Lee, Mark Hasegawa-Johnson, and Camille Goudeseune. Open loop multichannel inversion of room impulse response. *JASA*, 113(4):2202–2203, 2003.
- [69] Mohammed Kamal Omar and Mark Hasegawa-Johnson. Approximately independent factors of speech using nonlinear symplectic transformation. *IEEE Transactions on Speech and Audio Processing*, 11(6):660–671, 2003.

- [70] Mark Hasegawa-Johnson. Finding the best acoustic measurements for landmark-based speech recognition. *Accumu Magazine*, 11:45–47, 2002.
- [71] 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. *JASA*, 110(5):2961, 2001.
- [72] Mark Hasegawa-Johnson. Line spectral frequencies are the poles and zeros of a discrete matched-impedance vocal tract model. *Journal of the Acoustical Society of America*, 108(1):457–460, 2000.
- [73] Yanli Zheng and Mark Hasegawa-Johnson. Three dimensional tongue shape factor analysis. *ASHA Leader*, 5(16):144, 2000.
- [74] Mark Hasegawa-Johnson. Electromagnetic exposure safety of the carstens articulograph ag100. *Journal of the Acoustics Society of America*, 104:2529–2532, 1998.
- [75] 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. *JASA*, 102:3091, 1997.
- [76] Mark A. Hasegawa-Johnson. Burst spectral measures and formant frequencies can be used to accurately discriminate stop place of articulation. *JASA*, 98:2890, 1995.
- [77] 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.
- [78] Mark A. Johnson. A mapping between trainable generalized properties and the acoustic correlates of distinctive features. *JASA*, 94:1865, 1993.
- [79] 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.
- [80] Mark A. Johnson. Analysis of durational rhythms in two poems by robert frost. *MIT Speech Communication Group Working Papers*, 8:29–42, 1992.
- [81] Mark A. Johnson. Using beam elements to model the vocal fold length in breathy voicing. *JASA*, 91:2420–2421, 1992.

Manuscripts, Talks, and Student Theses

- [1] Heting Gao. Improving multilingual speech recognition systems. Master’s thesis, University of Illinois, 2021.
- [2] Junrui Ni. Enforcing constraints for multi-lingual and cross-lingual speech-to-text systems. Master’s thesis, University of Illinois, 2021.
- [3] Yijia Xu. Acoustic event, spoken keyword and emotional outburst detection. Master’s thesis, University of Illinois, 2019.
- [4] Amit Das. *Speech Recognition with Probabilistic Transcriptions and End-to-End Systems Using Deep Learning*. PhD thesis, University of Illinois, 2018.
- [5] 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.
- [6] Mary Pietrowicz. *Exposing the Hidden Vocal Channel: Analysis of Vocal Expression*. PhD thesis, University of Illinois, 2017.

- [7] Roger Serwy. *Hilbert Phase Methods for Glottal Activity Detection*. PhD thesis, University of Illinois, 2017.
- [8] Yang Zhang. *Generative Models for Speech and Time Domain Signals*. PhD thesis, University of Illinois, 2017.
- [9] Mark Hasegawa-Johnson. Limited data settings. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.
- [10] Mark Hasegawa-Johnson. Multivariate-state models for speech recognition. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.
- [11] Mark Hasegawa-Johnson. Prosody. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.
- [12] Mark Hasegawa-Johnson. Speech production, speech perception, and phonology. Lecture given at the Winter School on Speech and Audio Processing, Chennai, India, January 2016.
- [13] Mahmoud Abunasser. *Computational Measures of Linguistic Variation: A Study of Arabic Varieties*. PhD thesis, University of Illinois, 2015.
- [14] 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.
- [15] Austin Chen. Automatic classification of electronic music and speech/music audio content. Master’s thesis, University of Illinois, 2014.
- [16] Kaizhi Qian. Regularized estimation of gaussian mixture models for svm based speaker recognition. B.S. Thesis, University of Illinois, 2014.
- [17] Raymond Yeh. Divergence guided two beams viterbi algorithm on factorial hmms. B.S. Thesis, University of Illinois, 2014.
- [18] 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.
- [19] Mark Hasegawa-Johnson, David Harwath, Harsh Vardhan Sharma, and Po-Sen Huang. Transfer learning for multi-person and multi-dialect spoken language interface. 2012.
- [20] Jui-Ting Huang. *Semi-Supervised Learning for Acoustic and Prosodic Modeling in Speech Applications*. PhD thesis, University of Illinois, 2012.
- [21] Harsh Vardhan Sharma. *Acoustic Model Adaptation for Recognition of Dysarthric Speech*. PhD thesis, University of Illinois, 2012.
- [22] Xiaodan Zhuang. *Modeling Audio and Visual Cues for Real-world Event Detection*. PhD thesis, University of Illinois, April 2011.
- [23] 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.
- [24] Mark Hasegawa-Johnson, Jui-Ting Huang, Roxana Girju, Rehab Mustafa Mohammad Duwairi, Eiman Mohd Tayyeb H B Mustafawi, and Elabbas Benmamoun. Learning to recognize speech from a small number of labeled examples. Qatar Foundation Annual Research Forum, 2011.
- [25] Boon Pang Lim. *Computational Differences between Whispered and Non-whispered Speech*. PhD thesis, University of Illinois, 2011.

- [26] Jeremy Tiedemann. Characterization of the head-related transfer function using chirp and maximum length sequence excitation signals. Master’s thesis, University of Illinois, 2011.
- [27] Lae-Hoon Kim. *Statistical Model Based Multi-Microphone Speech Processing: Toward Overcoming Mismatch Problem*. PhD thesis, University of Illinois, August 2010.
- [28] Chi Hu. Fsm-based pronunciation modeling using articulatory phonological code. Master’s thesis, University of Illinois, 2010.
- [29] Arthur Kantor. *Pronunciation modeling for large vocabulary speech recognition*. PhD thesis, University of Illinois, 2010.
- [30] Sarah Borys. *Lovable Indestructible Grad Student of Chaos*. PhD thesis, University of Illinois, 2009.
- [31] Mark Hasegawa-Johnson. Pattern recognition in acoustic signal processing. Unpublished presentation at the Machine Learning Summer School, University of Chicago, 2009.
- [32] Bryce Lobdell. *Models of Human Phone Transcription in Noise Based on Intelligibility Predictors*. PhD thesis, University of Illinois, 2009.
- [33] Sarah Borys. Lovable indestructible grad student of chaos. Cartoons published online, 2008.
- [34] Sarah Borys. An svm front end landmark speech recognition system. Master’s thesis, University of Illinois, 2008.
- [35] Yang Li. Incremental training and growth of artificial neural networks. Master’s thesis, University of Illinois, 2008.
- [36] Harsh Vardhan Sharma. Universal access: Experiments in automatic recognition of dysarthric speech. Master’s thesis, University of Illinois, 2008.
- [37] Mark Hasegawa-Johnson. Audio-visual speech recognition: Audio noise, video noise, and pronunciation variability. talk given to the Signal Processing Society, IEEE Japan, June 2007.
- [38] Taejin Yoon. *A Predictive Model of Prosody Through Grammatical Interface: A Computational Approach*. PhD thesis, University of Illinois, 2007.
- [39] Mark Hasegawa-Johnson. Phonology and the art of automatic speech recognition. Director’s Seminar Series, Beckman Institute, University of Illinois at Urbana-Champaign, November 2006.
- [40] Mark Hasegawa-Johnson. Object tracking and asynchrony in audio-visual speech recognition. talk given to the Artificial Intelligence, Vision, and Robotics seminar series, August 2006.
- [41] Mark Hasegawa-Johnson. Dealing with acoustic noise. part i: Spectral estimation. July 2006.
- [42] Mark Hasegawa-Johnson. Dealing with acoustic noise. part ii: Beamforming. July 2006.
- [43] Camille Goudeseune and Bowon Lee. Avicar: Audio-visual speech recognition in a car environment. 2006.
- [44] Bowon Lee. *Robust Speech Recognition in a Car Using a Microphone Array*. PhD thesis, University of Illinois, 2006.
- [45] Rajiv Reddy. Analysis of pitch contours in repetition-disfluency using stem-ml. 2006.
- [46] Christopher Co and 2012. Room reconstruction and navigation using acoustically obtained room impulse responses and a mobile robot platform. Master’s thesis, University of Illinois, 2005.
- [47] Tae-Jin Yoon, Cole, Jennifer, Mark Hasegawa-Johnson, and Chilin Shih. Detecting non-modal phonation in telephone speech. unpublished manuscript, 2005.

- [48] Taejin Yoon. Mapping syntax and prosody. Presentation at the Midwest Computational Linguistics Colloquium, 2005.
- [49] Yanli Zheng. *Acoustic Modeling and Feature Extraction for Speech Recognition*. PhD thesis, University of Illinois, 2005.
- [50] Mark Hasegawa-Johnson. Landmark-based speech recognition: The marriage of high-dimensional machine learning techniques with modern linguistic representations. October 2004.
- [51] Mark Hasegawa-Johnson. Speech recognition models of the interdependence among syntax, prosody, and segmental acoustics. October 2004.
- [52] Ken Chen. *Prosody Dependent Speech Recognition on American Radio News Speech*. PhD thesis, University of Illinois, 2004.
- [53] Aaron Cohen. A survey of machine learning methods for predicting prosody in radio speech. Master's thesis, University of Illinois, 2004.
- [54] Stefan Geirhofer. Feature reduction with linear discriminant analysis and its performance on phoneme recognition. Undergraduate research project, 2004.
- [55] Sarah Borys. Recognition of prosodic factors and detection of landmarks for improvements to continuous speech recognition systems. B.S. Thesis, University of Illinois, 2003.
- [56] Ameya Deoras. A factorial hmm approach to robust isolated digit recognition in non-stationary noise. B.S. Thesis, University of Illinois, 2003.
- [57] Mohamed Kamal Mahmoud Omar. *Acoustic Feature Design for Speech Recognition: A Statistical Information-Theoretic Approach*. PhD thesis, University of Illinois, 2003.
- [58] Zhinian Jing. Voice index and frame index for recognition of digits in speech background. Master's thesis, University of Illinois, 2002.
- [59] David Petruncio. Evaluation of various features for music genre classification with hidden markov models. B.S. Thesis, University of Illinois, 2002.
- [60] Mark Hasegawa-Johnson. Preliminary work and proposed continuation: Imaging of speech anatomy and behavior. Unpublished presentation at Universities of Illinois Inter-campus Biomedical Imaging Forum, 2001.
- [61] Jul Setsu Cha. Articulatory speech synthesis of female and male talkers. Master's thesis, UCLA, 2000.
- [62] Wira Gunawan. Distributed speech recognition. Master's thesis, University of Illinois, 2000.
- [63] Mark A. Hasegawa-Johnson. *Formant and Burst Spectral Measurements with Quantitative Error Models for Speech Sound Classification*. PhD thesis, MIT, 1996.

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.

Tech Reports

- [1] David Cohen, Camille Goudeseune, and Mark Hasegawa-Johnson. Efficient simultaneous multi-scale computation of ffts. Technical Report GT-FODAVA-09-01, Georgia Institute of Technology, 2009.
- [2] 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. Technical Report WS06, Johns Hopkins Center for Language and Speech Processing, 2007.
- [3] 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. Technical Report WS04, Johns Hopkins Center for Language and Speech Processing, 2005.