

Carol Yvonne Espy-Wilson
Department of Electrical and Computer Engineering
Institute for Systems Research,
University of Maryland, College Park
espy@umd.edu

Education:

MIT, EECS, MS (1981) Eng. (1984) PhD (1987)

Stanford University, BSEE (1979)

Awards, Honors, and Special Recognition

- Keynote Speaker, 180th Meeting of the Acoustical Society of American, June 2021
- University of Maryland Campus Woman of Influence (2020)
- Keynote Speaker, 12th International Seminar on Speech Production, Dec. 2020
- Member, National Institute on Deafness and Communication Disorders Advisory Council, 2020-2023
- Member, National Institute on Biomed. Imaging and Bioengineering Advisory Council, 2015-2018
- Member, National Institute on Child Health and Human Development Advisory Board , 2010-2013
- First African American woman, and first African American, in ECE to achieve tenure and be promoted to the rank of full professor (University of Maryland First to ADVANCE Program, 2019)
- Jimmy Lin Award for Innovation in Electrical and Computer Engineering, University of Maryland, 2018

- Tutorial, Interspeech 2018, “Articulatory Representations: Measurement, Estimation and Application to State-of-the-Art Automatic Speech Recognition”
- Tutorial, 149th Meeting of the Acoustical Society of America, “Automatic Speech Recognition”
- Fellow, International Speech Community Association (ISCA), 2018
- Fellow, Radcliffe Institute for Advanced Study, Harvard University, 2008-2009
- Fellow of the Acoustical Society of America, Nov. 2005
- IEEE Senior Member
- Elected to the Speech and Language Technical Committee of IEEE, 2010-2012
- Elected Chair, Speech Technical Committee of ASA, Spring 2007-Fall 2009
- Member of the Advisory Council for the National Institutes on Deafness and Other Communication Disorders 2020-2023
- Member of the Advisory Council for the National Institutes on Biomedical Imaging and Bioengineering 2015-2018
- 2012-2013 Distinguished Scholar-Teacher Award, University of Maryland
- 2015-2017 Institute for Systems Research Senior Faculty Fellow Award
- Advance Professor 2011-2012
- Clare Boothe Luce Professor, August 1990 to August 1995
- Invention of the Year Award for “Multi-pitch tracking in adverse environments,” University of Maryland, 2010.
- 2010 Maryland Innovator of the Year Award for “Multi-pitch tracking in adverse environments”, Baltimore Daily Record.

- Featured in “*Stem Gems: How 44 Women Shine in Science, Technology, Engineering and Mathematics, And How You Can Too!*” published by Stephanie Espy in 2016
- First Place, University of Maryland \$75K Business Plan Competition, High Technology Category (\$25,000), 2010
- First Place, University of Maryland \$75K Business Plan Competition, Social Impact Category (\$15,000) , 2010
- StartRight! Women’s Business Plan Competition Award
- SAIC Award to a Venture Accelerator Company at UMD (OmniSpeech LLC)
- Honda Initiation Award, 2003 and 2004
- NIH Career Development Award, 1998-2003
- NIH Language and Communication Study Section, 2001-2004.
- NSF Research Initiation Award, 1990-1992
- Xerox Doctoral Fellow, 1981-1987

Editorial Service:

- Editorial Board, Computer Speech Language, 2021- present
- Associate Editor, Journal of the Acoustical Society of America, 2010-2018
- Editorial Board, Acoustics Today, Acoustical Society of America, 2007-2009

Academic Service:

University of Maryland College Park, Department of Electrical and Computer Engineering and Institute for Systems Research

2001-2007: Associate Professor, 2007-present: Full Professor

Boston University, Department of Electrical and Computer Engineering

1990-1999: Assistant Professor, 1999-2001: Associate Professor

Massachusetts Institute of Technology, Electrical Engineering and Computer Science

1988-1990: Research Scientist and Lecturer for 2 undergraduate courses; 1990-2000: Visiting Scientist; 1987-1988: Postdoctoral Fellow, Speech Communication Group; 1988-2001: Associate Housemaster in MacGregor Dormitory; 1985-1988: Graduate Tutor in MacGregor Dormitory

Industrial Experience

2011-present: Founder and CEO of OmniSpeech LLC

1987-1988: Part-time member of technical staff, Linguistics Research, Bell Telephone Laboratories, Murray Hill, New Jersey

Selected Invited Talks and Participations

- Invited lectures in academic and industrial research labs including MIT, Ga. Tech, NYU, Harvard, Stanford, UC Berkeley, University of Maryland Medical Division of Otolaryngology, National Institute of Mental Health, Brown University, Cirrus Logic, Nuance, Apple, John Hopkins, Northwestern University, Ohio State, Rutgers University, NIDCD workshop, MIT Lincoln Laboratories, Honda, Dartmouth College, Boston University, UCLA, Motorola
- Invited Speaker, Workshop on Computational Linguistics and Clinical Psychology, CLPsych 2021, Speech Acoustics for Mental Health Assessment”.
- Invited Speaker, From Sound to Sense Conference: 50+ Years of Discoveries in Speech Communication Conference, MIT, Cambridge, MA, June 11-13, 2004.
- Invited Speaker, National Institute of Mental Health’s Innovation Speaker Series, “Multimodal Signal Integration for Predictive Mental Health Analytics”, 2015.
- Invited Speaker, Symposia “Inspiring Women in Science”, Brown University, Celebration of Women’s History Month, a special one-day event with six women in science and engineering, 2017.
- Invited Speaker, NIH NIDCD workshop on Brain-Computer Interface Technologies, “Speech and Low Bandwidth Communication Technologies”, May 2006.

- Organized an NSF workshop on Speech for Robotics at University of Maryland, October 2019. 45 Participants from the US and abroad. Report written for program officers and is current in revision for Computer Speech and Language.
- Invited speaker on OmniSpeech LLC to the Committee on Economic Development and Technology Commercialization, Maryland board of Regents, Nov. 2013.
- Invited panelist, Science Coalition Panel for a Senate Staff Briefing, House Energy and Commerce Committee on Sparking Economic Growth, Capital Hill, Nov. 2013.
- External Evaluator, The Beckman Institute at the University of Illinois Urbana Champaign, Oct. 2015.
- Invited participant at several NSF panels and workshops (Directorates of Computer and Information Science and Engineering, and Education and Human Resources), and an NIH panel, 1995-present
- Keynote speaker for many events for students of color including some in the past 10 years: UMBC Meyerhoff Scholars Seminar (April 2019); Closing Reception for the Fall Regional NSBE Conference (Nov. 2018), Features Speaker in the STEM GEMs Virtual Summer Book Club (June, 2017), Keynote Speaker for Black Girls Who Code Event put on by Women in Engineering, Center for Minorities in Science and Engineering and DC Women of Color in STEM (Dec. 2016); College Board Conference (2015); MIT Summer Research Program (August 2012); Inaugural Willie Hobbs Moore Lecture Series, at U. Mich. Physics Department (March 2012); WEB DuBois Honor Society Induction Ceremony at UMD (2010); Center for Minorities in Science and Engineering/s Annual Student Recognition and Alumni Banquet (April 2008)

Refereed Journal Publications

27. M. Marge, C. Espy-Wilson, N. Ward, A. Alwan, Y. Artzi, M. Bansal, G. Blankenship, J. Chai, H. Daumé III, D. Dey , M. Harper, T.Howard, C. Kennington I. Kruijff-Korbayová, D. Manocha, C. Matuszek, R. Mead, R.Mooney, R. K. Moore, M. Ostendorf, H. Pon-Barry, A. Rudnický, M. Scheutz, R. St. Amant, T. Sun, S. Tellex, D. Traum, Z. Yu , “Spoken Language Interaction for Robotics: Research Issues and Recommendations”, accepted, Computer Speech Language, June 2021.

26. **S. Sahu**, R. Gupta and C. Espy-Wilson, “Modeling Feature Representations for Affective Speech using Generative Adversarial Networks”, accepted., IEEE Transactions on Affective Computing, May 2020.
25. **G. Sivaraman**, V. Mitra, H. Nam, M. Tiede, and C. Espy-Wilson, “Unsupervised speaker adaptation for speaker independent acoustic to articulatory speech inversion”, *The Journal of the Acoustical Society of America*, vol. 146, no. 1, 2019, pp. 316-329, doi: 10.1121/1.5116130.
24. V. Mitra, **G. Sivaraman**, H. Nam, C. Espy-Wilson, E. Saltzman, M. Tiede, (2017) “Hybrid Convolutional Neural Networks For Articulatory And Acoustic Information Based Speech Recognition”, *Speech Communication*, Volume 89, Issue C, Pages 103-112, May 2017.
23. S. Gordon-Salant, D. Zion and C. Espy-Wilson (2014) “Recognition of time-compressed speech does not predict recognition of natural fast-rate speech by older listeners” *Journal of the Acoustical Society of America*, Express Letters, vol. 136, pp. 268-274.
22. X. Zhou, J. Zhou, M. Stone, J. Prince and C. Espy-Wilson (2013), “Improve vocal tract reconstruction and modeling using an image super-resolution technique”, *Journal of the Acoustical Society of America Express Letters*, vol. 133, no. 6, pp. 439-445.
21. H. Nam, **V. Mitra**, M. Tiede, M. Hasegawa-Johnson, C. Espy-Wilson, E. Saltzman, L. Goldstein, (2012), “A procedure for estimating gestural scores from speech acoustics”, *Journal of the Acoustical Society of America*”, vol. 132, no. 6, 3980-3989.
20. **V. Mitra**, H. Nam, C. Espy-Wilson, E. Saltzman, and L. Goldstein (2012) “Recognizing articulatory gestures from speech for robust speech recognition”, *Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2270-2287.
19. **V. Mitra**, H. Nam, C. Espy-Wilson, E. Saltzman, L. Goldstein (2011), “Articulatory Information for Noise Robust Speech Recognition“, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 7, pp. 1913-1924.
18. **V. Mitra**, H. Nam, C. Espy-Wilson, E. Saltzman, L. Goldstein (2010) “Retrieving Tract Variables from Acoustics: a comparison of different Machine Learning strategies” *IEEE Journal of Selected Topics*, vol. 4, issue 6, pp. 1027-1045.

17. **X. Zhou**, C. Espy-Wilson, S. Boyce, M. Tiede, Christy Holland and Ann Choe (2008) “A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/ sounds”, *Journal of the Acoustical Society of America*, Vol. 123, no. 6, pp. 4466-4481.
16. **A. Juneja** and Carol Espy-Wilson (2008) “Probabilistic landmark detection for automatic speech recognition using acoustic-phonetic information”, *Journal of the Acoustical Society of America*, vol. 123, no. 2, pp. 1154-1168.
15. **T. Pruthi**, C. Espy-Wilson and Brad Story (2007) “Simulation and analysis of nasalized vowels based on MRI data”, *Journal of the Acoustical Society of America*, vol. 121, no. 6, pp. 3858-3873.
14. **O. Deshmukh**, C. Espy-Wilson, L. Carney, (2007) “Speech Enhancement using the Modified Phase Opponency Model”, *Journal of the Acoustical Society of America*, vol. 121, no. 6, pp. 3886-3898.
13. **O. Deshmukh**, C. Espy-Wilson, **A. Salomon**, and **Jawahar Singh** (2005) “Use of Temporal Information: Detection of the Periodicity, Aperiodicity and Pitch in Speech”, *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 776-786.
12. **T. Pruthi** and C. Espy-Wilson (2004) “Acoustic Parameters for Automatic Detection of Nasal Manner”, *Speech Communication*, vol. 43, no. 3, pp.225-239.
11. **Z. Zhang** and Carol Espy-Wilson (2004) “A vocal tract model for American English /l/”, *Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1274-1280.
10. **A. Salomon**, C. Espy-Wilson and Om Deshmukh (2004) “Detection of Speech Landmarks: Use of Temporal Information”, *Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1296-1305.
9. M. Jackson, C. Espy-Wilson and S. Boyce (2001) “Verifying a vocal tract model with a closed side branch”, *Journal of the Acoustical Society of America*, vol. 109, no. 6, pp. 2983-2987.
8. C. Y. Espy-Wilson, S. Boyce, M. Jackson, S. Narayanan and A. Alwan (2000) “Acoustic Modeling of American English /r/”, *Journal of the Acoustical Society of America*, pp. 343-356.
7. F. Guenther, C. Espy-Wilson, S. Boyce, M. Matthies, M. Zandipour and J. Perkell (1999) “Articulatory tradeoffs reduce acoustic variability during American English /r/

production”, *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2854-2865.

6. C. Y. Espy-Wilson, V. R. Chari, J. M. MacAuslan, C. B. Huang and M. J. Walsh (1998) “Enhancement of Electrolaryngeal Speech by Adaptive Filtering”, *Journal of Speech, Language and Hearing Research*, vol. 41, no. 6, December, pp. 1253-1264.
5. S. Boyce and C. Y. Espy-Wilson (1997) “Coarticulatory Stability in American English /r/s”, *Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3741-3753.
4. V. Chari and C. Espy-Wilson (1995) “Extraction of Formant Frequencies by Adaptive Enhancement of Fourier Spectra,” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 35-39.
3. C. Y. Espy-Wilson (1994) “A Feature-Based Semivowel Recognition System,” *Journal of the Acoustical Society of America*, vol. 96, no. 1, pp. 65-72.
2. C. Y. Espy-Wilson (1992) “Acoustic Measures for Linguistic Features Distinguishing the Semivowels /wjr/ in American English,” *Journal of the Acoustical Society of America*, vol. 92, no. 2, pp. 736-751.
1. C. Y. Espy and J. Lim (1983) “Effects of Additive Noise on Signal Reconstruction from Fourier Transform Phase,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 894-898.

Book Chapters

1. C. Espy-Wilson, G. Sivaraman, M. Tiede, V. Mitra, E. Saltzmann, L. Goldstein, H. Nam, “Modeling of Articulatory Gestures to Control Effects of Production Variability on Speech Technologies”. In Cangemi, Clayards, Niebuhr, Schupler & Zellers (eds). *Rethinking Reduction*, Berlin: Mouton de Gruyter, 2018.
2. M. Tiede, S. Boyce and C. Espy-Wilson, “Variability of North-American /r/ production in response to palatal perturbation”. In Maassen, Ben and van Lieshout, Pascal (eds.), *Speech Motor Control: New developments in basic and applied research*, Oxford University Press, 2010.
3. Carol Espy-Wilson, “Phonological Models of Variation in Computer Speech Processing: Commentary on the papers by Nam, Son et al., and Hirschberg”. In Cole, Jennifer and Hualde, José I. (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter. pp. 535-546,

2007.

4. S. Nawab, C. Espy-Wilson, R. Mani, and N. Bitar, "Knowledge-Based Analysis of Speech Mixed with Sporadic Environmental Sounds," Computational Auditory Scene Analysis edited by Rosenthal and Okuno Lawrence Erlbaum Associates Inc. Publishers, 1998.

Selected Peer-Reviewed Conference Proceedings (last 10 years)

1. N. Seneviratne, C. Espy-Wilson, "Generalized Dilated CNN Models for Depression Detection Using Inverted Vocal Tract Variables", accepted for Interspeech 2021, Aug. 30-Sept. 3, 2021 (arxiv: 2011.06739)
2. N. Seneviratne, C. Espy-Wilson, "Speech based Depression Severity Level Classification Using a Multi-Stage Dilated CNN-LSTM Model", accepted for Interspeech 2021, Aug. 30-Sept. 3, 2021 (arxiv: 2104.04195)
3. Maduwantha, Y., Kitchen, C., Kelly, D.L. & Espy-Wilson, C. (2021). Inverted vocal tract variables and facial action units to quantify neuromotor coordination in schizophrenia. In Proceedings of the 12th International Seminar on Speech Production (ISSP 2020), 174-177.
4. N. Seneviratne, A. Lammert, Jim Williamson, T. Quatieri and C. Espy-Wilson "Extended Study on the Use of Vocal Tract Variables to Quantify Neuromotor Coordination in Depression", Proc. Interspeech 2020.
5. S. Sahu, V. Mitra, N. Seneviratne and C. Espy-Wilson. "Multi-Modal Learning for Speech Emotion Recognition: An Analysis and Comparison of ASR Outputs with Ground Truth Transcription", Proc. of Interspeech, Sept. 2019.
6. N. Seneviratne, G. Sivaraman, V. Mitra and C. Espy-Wilson, "Noise Robust Articulatory to Acoustic Speech Inversion, Proc. of Interspeech, Sept. 2019.
7. C. Espy-Wilson, A. Lammert, N. Seneviratne, T. Quatieri, "Assessing Neuromotor Coordination in Depression Using Inverted Vocal Tract Variables", Proc. Interspeech 2019, 1448-1452, DOI: 10.21437/Interspeech.2019-1815.
8. S. Sahu and C. Espy-Wilson, "On Enhancing Speech Emotion Recognition Using Generative Adversarial Networks", Proc. of Interspeech 2018.
9. N. Seneviratne, G. Sivaraman, V. Mitra and C. Espy-Wilson, "Noise Robust Articulatory to Acoustic Speech Inversion, Proc. of Interspeech 2018.

10. S. Sahu, R. Gupta, G. Sivaraman, and C. Espy-Wilson, "Smoothing Model predictions Using Adversarial Training Procedures for Speech based Emotion Recognition," Proc. of ICASSP 2018.
11. R. Gupta, S. Sahu, C. Espy-Wilson and S. Naryanan, "Semi-supervised and Transfer learning approaches for low resource sentiment classification," Proc. of ICASSP 2018.
12. Mitra, V., Sivaraman, G., Bartels, C., Nam, H., Wang, W., Espy-Wilson, C., Vergyri, D. and Franco, H., 2017, March. Joint modeling of articulatory and acoustic spaces for continuous speech recognition tasks. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5205-5209). IEEE.
13. R. Gupta, S. Sahu, C. Espy-Wilson, and S. Narayanan. An Affect Prediction Approach through Depression Severity Parameter Incorporation in Neural Networks. In In Proceedings of Interspeech, August 2017.
14. S. Sahu, R. Gupta, G. Sivaraman, W. AbdAlmageed, and C. Espy-Wilson, "Adversarial auto-encoders for speech based emotion recognition," in Interspeech, 2017, pp. 1243–1247.
15. G. Sivaraman, V.Mitra, H. Nam, M.K. Tiede, C. Espy-Wilson (2016) Vocal tract length normalization for speaker independent acoustic-to-articulatory speech inversion, Proc. of INTERSPEECH 2016.
16. S.Sahu, C.Espy-Wilson, "Speech features for depression detection", In Proceedings of Interspeech Sept. 2016, pp. 1928-1932.
17. V. Mitra, W. Wang, Y. Lei, A. Kathol, G. Sivaraman and C. Espy-Wilson, "Robust features and system fusion for reverberation and their role in speech recognition", ICASSP, 2015.
18. G. Sivaraman, V. Mitra, M. Tiede, E. Saltzman, L. Goldstein and C. Espy-Wilson, "Analysis of coarticulated speech using estimated articulatory trajectories", Interspeech, 2015.
19. V. Mitra, G. Sivaraman, H. Nam, C. Espy-Wilson, E. Saltzman, "Articulatory features from deep neural networks and their role in speech recognition" Proceedings of ICASSP 2014.
20. D. Zion, C. Espy-Wilson, S. Gordon-Salant, "Recognition of natural-rate, time-compressed, and natural fast-rate sentences by younger and older listeners", presented at the Aging and Speech Communication Conference, 5th International and Interdisciplinary Research Conference, Indiana University, Bloomington, October 6-9, 2013.

21. G. Sivaraman, V. Mitra and C. Espy-Wilson, "Fusion Of Acoustic, Perceptual And Production Features For Robust Speech Recognition In Highly Non-Stationary Noise" presented at the CHIME Challenge, ICASSP 2013.
22. X. Zhou, J. Woo, M. Stone and C. Espy-Wilson, "A Cine Mri-Based Study Of Sibilant Fricatives Production In Post-Glossectomy Speakers", Proceedings of ICASSP 2013, pp. 7780-7784.
23. X. Zhou, D. Garcia-Romero, N. Mesgarani, M. Stone, C. Espy-Wilson and S. Shamma "Automatic intelligibility assessment of pathologic speech in head and neck cancer based on auditory-inspired spectro-temporal modulations", Proc. of Interspeech, September 2012.
24. D. Garcia-Romero, X. Zhou and C. Espy-Wilson, "Multicondition Training of Gaussian PLDA Models In I-Vector Space For Noise And Reverberation Robust Speaker Recognition, Proceedings of ICASSP 2012.
25. X. Zhou, D. Garcia-Romero, R. Duraiswami, C. Espy-Wilson and S. Shamma, "Linear versus Mel- Frequency Cepstral Coefficients for Speaker Recognition" Automatic Speech Recognition and Understanding Workshop, Dec. 2011, Hawaii.
26. V. Mitra, H. Nam and C. Espy-Wilson, "Robust speech recognition using articulatory gestures in a Dynamic Bayesian Network framework, Automatic Speech Recognition and Understanding Workshop, Dec. 2011, Hawaii.
27. J. Zhou, D. Garcia-Romero, and C. Espy-Wilson, "Automatic Speech Codec Identification with Applications to Tampering Detection of Speech Recordings", Proceedings of Interspeech 2011, pp. 2533-2536.
28. D. Garcia-Romero and C. Espy-Wilson, "Analysis of I-vector Length Normalization in Speaker Recognition Systems", Interspeech 2011, pp. 249-252.
29. X. Zhou, M. Stone and C. Espy-Wilson, "A comparative acoustic study on speech of glossectomy patients and normal subjects", Interspeech 2011, pp.517-520.
30. V. Mitra, H. Nam, C. Espy-Wilson, E. Saltzman, L. Goldstein, "Speech Inversion: Benefits of Tract Variables over Pellet Trajectories", Proceedings of ICASSP 2011.
31. V. Mitra, H. Nam, C. Espy-Wilson, E. Saltzman, L. Goldstein, "Gesture-based Dynamic Bayesian Network for Noise robust Speech Recognition" , Proceedings of ICASSP 2011.

32. V. Mahadevan and C. Espy-Wilson, "Maximum Likelihood Pitch Estimation Using Sinusoidal Modeling, Proceedings of the International Conference on Communications and Signal Processing, 2011.
33. H. Nam, V. Mitra, M. Tiede, E. Saltzman, L. Goldstein, C. Espy-Wilson, M. Hasegawa-Johnson, A procedure for estimating gestural scores from natural speech, Proc. of Interspeech, pp. 30-33, Japan, 2010.
34. V. Mitra, H. Nam, C. Espy-Wilson, E. Saltzman, L. Goldstein, Robust Word Recognition using articulatory trajectories and Gestures, Proc. of Interspeech, pp. 2038-2041 Japan, 2010.
35. X. Zhou, C.Y. Espy-Wilson, M. Tiede, S. Boyce, "An MRI-based articulatory and acoustic study of lateral sound in American English", Proceedings of ICASSP 2010.
36. D. Garcia-Romero and C. Espy-Wilson, "Automatic Acquisition Device Identification from Speech Recordings" Proceedings of ICASSP 2010.

Supervised: 10 PhD, 31 Master's, and 40 undergraduate students

Academic Research Program

- **Grants and Contracts:** More than 7 million dollars from NSF, NIH, DoD, industry.
- **Patents:** 7

OmniSpeech LLC

- **Grants:** More than 1 million dollars from NSF (SBIR Phase, 1, Phase II and Phase IIB) and NIH (STTR).
- **Angel Investments:** More than 5 million dollars.
- **Patents:** 6 and another 3 pending