Meghavarshini Krishnaswamy

Research Statement

My research experience includes phonetic accommodation, L2 speech acquisition, pitch perception and its effects on phonemic identification, and the acoustics of complex tongue gestures. I have worked on the sound systems of Indic languages such as Malayalam, Bangla, Hindi, and Assamese on projects investigating the phonetic realisation of phonetic contrast in languages with dense consonant inventories. During my PhD, I expanded my experience in phonological phonetics and laboratory phonology to include phonetics and speech technology, in order to study the acoustic aspects of team behaviour and interaction. My dissertation research investigates the linguistic aspects of vocal entrainment and phonetic accommodation, to find a relationship between convergence in speech patterns and cooperation among speakers in a team setting.

Besides phonetics, my research interests include internet linguistics, Englishes of the world, and NLP applications for protected data.

Deep-learning applications for multi-party vocal entrainment

A question that has motivated my research in linguistics is: How do humans use their linguistic apparatus to impact not just what they say, but how they say it? In a social setting, conversations involve the exchange of information, as well as the sending and receiving of cues for building rapport and signalling cohesion. This entrainment (also researched under the terms 'coordination', 'synchrony' and 'accommodation') has been the subject of interdisciplinary research in psychology, linguistics, SLP, cognitive and behavioural sciences, and anthropology. Within experimental linguistics, entrainment has been reported across multiple linguistic levels such as the syntactic (Branigan et al., 2000), lexical (Ward & Litman, 2007) and phonetic (Babel, 2009; Borrie et al., 2019; Litman et al., 2016), and across multiple languages such as Russian, Hebrew, Slovak and Mandarin (Beňuš et al., 2014; Kachkovskaia et al., 2020; Levitan et al., 2015; Weise et al., 2020). I am fascinated by the rich history of studying human behaviour using linguistic data, and its impact on understanding the outcomes for tasks that require cohesion and good communication, such as psychological counselling (Nasir et al., 2017), conversation with strangers (Nasir et al., 2020), team activities (Beňuš, 2014; Litman et al., 2016), and group projects (Friedberg et al., 2012).

My research (funded by the ToMCAT Project) focuses on expanding our understanding of entrainment in two main ways- building experiments to study entrainment in non-dyadic (or multi-party) conversations among strangers engaged in a team task, and studying the efficacy of modelling vocal entrainment with unsupervised deep learning models. The first objective is challenging and necessary, as intimacy and balanced turn-taking cannot be readily assumed in two-party conversations. The second contribution attempts to understand if the research on entrainment can be scaled up to include larger datasets and more complex acoustic feature sets for the modelling.

Given the historical difficulty in data processing, statistical modelling and annotating for entrainment, I chose to explore the viability of deep neural networks and their ability to learn patterns from unstructured data. I examine if entrainment can be better understood with a different methodology. Another motivation for my doctoral research is to find out if linguistic entrainment can be used to improve the study of multi-modal team behaviour and communication, which has many real-world applications. My work on multi-party entrainment has resulted in two publications (proceedings): "The ToMCAT Dataset" (Pyarelal et al., 2024) and "Probabilistic Modeling of Interpersonal Coordination Processes" (Soares et al., 2024).

My research experience and training, in combination with my work as an NLP consultant and workshop lead at the Data Science Institute (DSI) of the University of Arizona, has assisted me in making connections between theoretical aspects of computational linguistics, and its importance in studying AI applications in healthcare, collegiate teaching, and human subject data processing.

Post-doctoral Research Plan

The Maryland Language Science Center is a great fit for exploring my future research because it is connected to many researchers and labs that explore issues in social sciences with a focus on speech data, which is a great fit for exploring the correlates of group cohesion in speech and multimodal communication. For this postdoc, I am interested in expanding the scope of my research in computational linguistics and speech

mkrishnaswamy@arizona.edu • meghavarshini.github.io
https://github.com/meghavarshini

technology, to further investigate multi-party entrainment in a broader range of speech settings. I intend to add to research on vocal entrainment on three fronts: the effect of interlocutors' sociolinguistic background on the directionality and quality of entrainment, the impact of signal quality and environmental noise on the robustness of the findings, and the push-pull relationship between variability and diversity in human speech and the natural process of converging on speech patterns during interactions.

First, while entrainment has been understood extensively at the language level across multiple languages, I want to explore the impact of dialect, social hierarchy and age on the directionality of entrainment and speech cohesion, and assess if automatic methods of detecting entrainment capture them. This work intersects with the research projects lead by Professor Charlotte Vaughn, whose sociolinguistic focus and expertise with speech data among different racial and ethnic populations would provide me with a direction I am keen to explore. Beyond human-human communication, I want to extend this research to examine multi-party entrainment during human-AI interactions with both L1 and L2 speech, to better understand the directionality of entrainment when one of the interlocutors is not human and the variety used is non-standard. This could add to the ongoing research on AI agents in team settings, and have an impact on how to simulate naturalness in human-AI interactions, going beyond one-way (non-interactive) settings.

I am also interested in directionality of entrainment when the interlocutors have a hierarchical relationship. In my dissertation I have studied entrainment between interlocutors who are strangers and are playing a collaborative video game. They have no power relationship to each other. I want to consider the impact of social hierarchies such as boss-employee, teacher-student, leader-follower, and dialectal prestige on the direction of entrainment. Research on vocal entrainment can be extended in the future to examine such sociolinguistic variables that reflect prestige and social power. Since my current method allows for studying this topic across multiple languages and dialects, it would be suitable for studying social relationships in many cultural settings.

A big confound for research with speech data is the quality of the recorded data. For my dissertation research, a large portion of the data collection was done remotely, with diverse recording conditions, equipment, and background noise. As a second avenue, I want evaluate how this variation impacts the perception and detection of entrainment. While learning more about the work done at the Maryland Language Science Center, I was fortunate to come across the research of Professor Rochelle Newman, whose focus on ASR and noise would provide a great avenue for understanding if people feel connected to one another at different levels of environmental noise, and how entrainment strategies vary when environmental conditions are different.

Thirdly, I want to outline a methodology for quantifying how much a speaker can diverge from their individual speaker characteristics to converge with their interlocutors. Do speakers with a greater deal of accent and voice quality similarities demonstrate higher levels of entrainment, and is this observed in the actual data? Or, does the higher inherent similarity in linguistic patterns among interlocutors make it more difficult to observe convergence when they interact with one another? Does entrainment among speakers with higher degrees of accent/cultural differences increase or decrease the degree to which teammates converge after a period of interaction, and does it make it easier or more difficult to quantify the changes in vocal characteristics? How do the interlocutors' mental states impact this balance between individual variation and an unconscious drift towards a group speech pattern? Given the strong relationship between team cohesion and mental state, I would be grateful for collaboration in experimental design and data collection methodology from researchers like Professor Naomi Feldman, whose experience in computational language modelling and the effect of mental states (such as feeling of comfort and positively with new team mates) on communication could contribute immensely to modelling the psychological correlates of multi-party entrainment.

References

Babel, M. E. (2009). *Phonetic and social selectivity in speech accommodation*. University of California, Berkeley.

Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation*, *6*, 802–813.

Beňuš, Š., Levitan, R., Hirschberg, J., Gravano, A., & Darjaa, S. (2014). Entrainment in Slovak collaborative dialogues. *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*, 309–313.

Borrie, S. A., Barrett, T. S., Willi, M. M., & Berisha, V. (2019). Syncing up for a good conversation: A clinically meaningful methodology for capturing conversational entrainment in the speech domain.

Journal of Speech, Language, and Hearing Research, *62*(2), 283–296. https://doi.org/10.1044/2018_JSLHR-S-18-0210

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25.

Friedberg, H., Litman, D., & Paletz, S. B. (2012). Lexical entrainment and success in student engineering groups. *2012 IEEE spoken language technology workshop (SLT)*, 404–409.

Kachkovskaia, T., Chukaeva, T., Evdokimova, V., Kholiavin, P., Kriakina, N., Kocharov, D., Mamushina, A., Menshikova, A., & Zimina, S. (2020). SibLing corpus of Russian dialogue speech designed for research on speech entrainment. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 6556–6561. https://aclanthology.org/2020.lrec-1.807

Levitan, R., Beňuš, Š., Gravano, A., & Hirschberg, J. (2015). Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 325–334.

Litman, D., Paletz, S., Rahimi, Z., Allegretti, S., & Rice, C. (2016). The teams corpus and entrainment in multi-party spoken dialogues. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1421–1431.

Nasir, M., Baucom, B., Bryan, C., Narayanan, S., & Georgiou, P. (2020). Modeling vocal entrainment in conversational speech using deep unsupervised learning. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2020.3024972

Nasir, M., Baucom, B. R., Bryan, C. J., Narayanan, S. S., & Georgiou, P. G. (2017). Complexity in speech and its relation to emotional bond in therapist-patient interactions during suicide risk assessment interviews. *Interspeech*, 3296–3300.

Pyarelal, A., Duong, E., Shibu, C., Soares, P., Boyd, S., Khosla, P., Pfeifer, V. A., Zhang, D., Andrews, E., Champlin, R., Raymond, V., Krishnaswamy, M., Morrison, C., Butler, E., & Barnard, K. (2024). The ToMCAT dataset. *Advances in Neural Information Processing Systems*, *36*. %5Curl%7Bhttps://proceedings.neurips.cc/paper_files/paper/2023/file/803d8d4b4a549d0d062fc704f8659ce3-Paper-Datasets_and_Benchmarks.pdf%7D

Soares, P., Pyarelal, A., Krishnaswamy, M., Butler, E., & Barnard, K. (2024). Probabilistic modeling of interpersonal coordination processes. *Forty-first International Conference on Machine Learning*.

Ward, A., & Litman, D. J. (2007). Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. *SLaTE Workshop on Speech and Language Technology in Education*. http://d-scholarship.pitt.edu/23210/

Weise, A., Silber-Varod, V., Lerner, A., Hirschberg, J., & Levitan, R. (2020). Entrainment in spoken Hebrew dialogues. *Journal of Phonetics*, *83*, 101005.