

### Overview

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. During my PhD, I expanded my experience in phonological phonetics and laboratory phonology to include 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.

### Acoustic correlates of phonemic contrast in dense inventories

While my dissertation focuses on computational linguistics and the study of entrainment, I also have a strong background in phonological phonetics. My main interest in phonetics has been in investigating phonemic contrast among Indic languages, and the impact of resistance to coarticulation on the perception and access of different phonemes. This research explored articulatory, acoustic, and perceptual studies during my pre-

doctoral and doctoral studies.

During my time on the project, “Influence of coarticulation on lexical access” (DST-CSRI, Govt. of India), we assessed the differences in coarticulation patterns among Malayalam dental, alveolar, and retroflex geminate stops with their neighbouring vowels, and examined the relationship between coarticulatory effects, lingual gestural complexity, and lexical frequency and for three sounds with very close places of articulation. My main role was designing the stimuli inventory and word lists for all experiments, conducting the production experiments, and constructing datasets, and writing code for the data analysis.

Our speech production study found that coarticulation trends with both the preceding and succeeding vowels, defied the predictions based on the relatively higher gestural complexity for retroflexes and dentals (Bladon & Al-Bamerni, 1976; Recasens et al., 1997), but could be explained by the impact of the lower and sparser availability of alveolar stops in the Malayalam. This led us to perform a subsequent ultrasound study to investigate the tongue shapes involved in the production of this three-way contrast, whose findings correlated with the acoustic study. A perceptual study of Malayalam listeners (GP et al., 2023) showed compared to unaltered recordings, stimuli with mismatched coarticulatory information (that is, a given coronal stop replaced with a different one) resulted in higher confusibility, further highlighting the role coarticulation plays in lexical access of tight contrasts.

My findings from a perception study with listeners who do not have this contrast and the associated lexical access needs (native listeners of American English) perceive these sounds that are unusually articulatorily close (Krishnaswamy & Warner, 2023). American English speakers have a high degree of familiarity with alveolars but low familiarity with dentals and retroflexes. We found that after a brief training period, listeners were more able to distinguish retroflexes from the other two categories, which adds to our understanding of how differences in complexity of the distinction impacts perception.

My phonetics research also includes work on gemination in Bengali through an acoustic study (Krishnaswamy et al., 2018). This work investigated how voicing continues despite aerodynamic constraints during the voiced geminate closure, and how Bengali speakers adjust tongue shape in a way that facilitates voicing continuing through the closure. Bengali is an ideal language to study this topic because it has both length and voicing distinctions in stops.

## Future Research

I am interested in expanding the scope of my research in computational linguistics and speech technology, to further investigate multi-party entrainment in a broader range of speech settings. I am interested in looking at the effect of interlocutors' L2 background (and linguistic prestige of different accents) on the directionality and quality of entrainment, as well as its impact on the cooperative task itself. I would particularly like 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, so they have no power relationship to each other. In future research, it would be useful to consider the impact of social hierarchies such as boss-employee, teacher-student, leader-follower, and dialectal prestige on the direction of entrainment. This could be extended in the future to examine sociolinguistic variables that reflect prestige and social power. My current method allows for studying this topic in any language, which provides opportunities for studying social relationships in many cultures.

For future research in phonetics, I would like to examine how the complexity of phonemic inventories impacts articulation, acoustics, and perception in a wider variety of languages. My work thus far has built methods that can be applied to a wide range of other languages and language families with different complexity to their phonemic inventories, such as Austronesian coronal stop contrasts and varieties of Arabic with voiced geminate stops. The work I did on Bengali geminates is already leading to potential future collaborations on Arabic. I would also like to see if my findings on the acoustics of human speech map onto how ASR systems classify distinctions such as the three-way contrast in Malayalam. This would allow investigation of whether the human and computational perceptual systems are using related methods.

## 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.
- Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulation resistance in English/I. *Journal of Phonetics*, 4(2), 137–150.
- 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](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.
- GP, S., Krishnaswamy, M., Mishra, R., & Dutta, I. (2023). Mismatched coarticulatory information hinders lexical access of coronal stops in Malayalam. *Proceedings of the 20th International Congress of Phonetic Sciences*, 371–375. [https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full\\_papers/657.pdf](https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/657.pdf)
- 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>
- Krishnaswamy, M., Dutta, I., & Banerjee, U. (2018). Active cavity expansion through lingual adjustments to place of constriction in voiced geminates. *Proceedings of Meetings on Acoustics*, 33(1), 060002. <https://asa.scitation.org/doi/pdf/10.1121/2.0001024>
- Krishnaswamy, M., & Warner, N. (2023). Perception of Malayalam three-way stop contrast among American English speakers. *Proceedings of the 20th International Congress of Phonetic Sciences*, 401–405. [https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full\\_papers/682.pdf](https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2023/full_papers/682.pdf)
- 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., Allegritti, 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. [https://proceedings.neurips.cc/paper\\_files/paper/2023/file/803d8d4b4a549d0d062fc704f8659ce3-Paper-Datasets\\_and\\_Benchmarks.pdf%7D](https://proceedings.neurips.cc/paper_files/paper/2023/file/803d8d4b4a549d0d062fc704f8659ce3-Paper-Datasets_and_Benchmarks.pdf%7D)
- Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America*, 102(1), 544–561.
- 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.