Commentary on paper [P1] Asif Salekin

[P1] Harshit Sharma, Yi Xiao, Victoria Tumanova, **Asif Salekin**, "Psychophysiological Arousal in Young Children Who Stutter: An Interpretable AI Approach", Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), 2022. https://doi.org/10.1145/3550326

Problem and Motivation: Speech production is a complex process requiring precise vocal tract coordination while simultaneously processing cognitive-linguistic information. Naturally, speech production can be affected by the speaker's physiological arousal. Literature shows that young children who stutter (CWS) are especially vulnerable to such influences. Given that preschool age is the time when essential communication skills are undergoing the most significant development, and also when some children develop stuttering, it is essential for our understanding of stuttering to examine young children's physiological responses during speech production. *This study presents an interpretable AI approach to identify the second-by-second fluctuations and pattern differences in physiological arousal of preschool-age children who stutter (CWS) compared to others who don't during various speaking tasks.*

Paper's Novelty and Contribution: The challenges of the paper include a lack of data annotations since the subtle physiological markers of CWS are yet unknown. However, we know that the data belong to CWS or others, making the data labels weakly labeled. Our preliminary analysis demonstrated that the latent patterns indicative of CWS's unique situational response from different physiological sensors, such as heart rate (HR), EDA, and respiratory rate (RSP), are sparse and do not emerge simultaneously; hence, there is a need for modality-wise distinctive pattern identification. However, literature has shown that cross-modality integration of these sensors is also highly effective in affective state and physiological arousal assessment.

Hence, we developed an approach, MI-MIL, that identifies the CWS's distinctive sparse patterns independently from each physiological modality without any available annotations of such patterns and captures and leverages the cross-relationships of the identified modality-specific sparse patterns for effective CWS vs. CWNS classification. To address the weakly labeled data challenge and identify modality-wise distinct patterns, MI-MIL applies the modality-wise multiple-instance-learning (MIL) paradigm in each physiological modality independently. MIL paradigm is designed to extract sparse and subtle patterns from weakly labeled data (i.e., without any fine-grain annotations of the region, timestamps, or duration of the patterns in the data). To capture the cross-modality-relations, MI-MIL presents a novel modality-fusion network that identifies the cross-relations of each modality's CWS indicative sparse patterns.

Moreover, leveraging SHAP, we visualize and discuss the fine-grain, second-by-second, temporal, and distinctive physiological response (represented through physiological parameters: HR, EDA, RSP) patterns of CWS from others during speech production in different challenging conditions. This paper is the first to leverage and discuss machine learning model interpretation and its findings on stuttering individuals. Such visualization and identification of patterns have both group-wise and personalized impacts. Identifying and visualizing group-wise patterns would enhance our understanding of stuttering etiology and development that eventually can improve clinical services for people who stutter. Personalized temporal pattern identification would enable remote, continuous, and real-time monitoring of stuttering children's physiological arousal, which may lead to personalized, just-in-time interventions to mitigate their arousal responses during speaking, resulting in an improvement in speech fluency.

Professional Significance: The paper published in csranking.org listed top conference Ubicomp/IMWUT 2022. The work resulted from my lab's ongoing collaboration with the Syracuse University Speech clinic, funded by two NIH grants (#$R21DC018103$ and # $R01DC017476-S2$). My Ph.D. student Harshit Sharma is the first author, my Ph.D. student Yi Xiao is the second author, and I am the corresponding author.