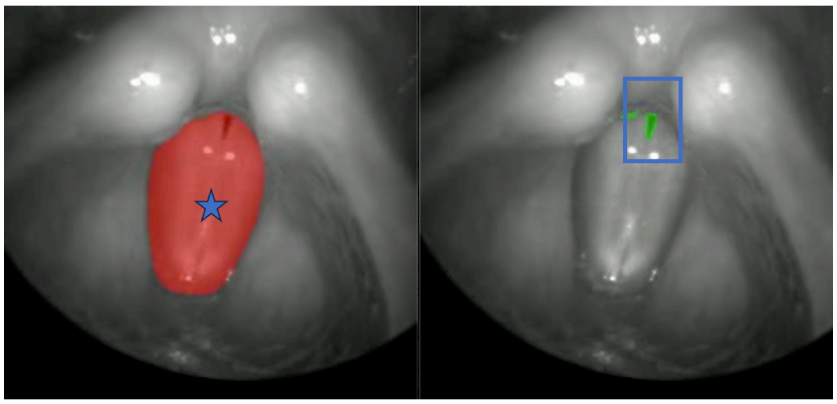
BSc thesis

Assessing foundation models in laryngeal endoscopy

Background. Laryngeal endoscopy, especially high-speed videoendoscopy (HSV), is a crucial tool in the assessment of vocal fold function and overall laryngeal health. Unlike standard endoscopic imaging, HSV captures high-frame-rate videos that allow detailed analysis of rapid vocal fold vibrations, providing insights that are valuable for diagnosing and treating voice disorders. Segmenting the vocal folds and the glottal area (the space between the vocal folds) in these videos is essential for quantitative analysis. Precise segmentation enables the extraction of key biomechanical parameters, such as glottal width, vibratory symmetry, and open/closed phase ratios. These parameters are vital for assessing vocal fold motion, identifying abnormal vibratory patterns, and tailoring therapeutic interventions for individuals with vocal pathologies. Accurate segmentation thus facilitates reliable, objective metrics that contribute to enhanced diagnostic precision and treatment outcomes. However, deep neural networks that work in all cases out-of-the-box are hard to achieve. Segment Anything-like models with little user input seem to be a promising alternative.



Project idea. In this project, you will assess foundation models (especially SAMv2) in a laryngeal endoscopy setting. You will identify which user input is best suited for temporal segmentation of endoscopy footage and how these foundation model-generated segmentations compare to fully supervised trained deep neural networks.

Work package (WP) 1: Use existing annotations or create new annotations for the BAGLS dataset as a ground truth for your experiments
WP 2: Use SAMv2 with point and box annotations for the vocal folds and the glottal area to generate semantic segmentation masks. Train a deep neural network (U-Net variant) segmenting the vocal folds and/or the glottal area.
WP 3: Compute quality metrics and evaluate how foundation models compare to specialized deep neural networks trained fully supervised on this task.

Your background. You are proficient in Python, numpy, scipy and matplotlib. You would like to get in touch with latest generative AI for medical image analysis. You are open to hands-on projects and eager to produce solid results. We believe this project is ideally suited for a BSc thesis (medical engineering, computer science) and could lead to a peer-reviewed publication.

Contact: Prof. Dr. Andreas Kist, Juniorprofessur für Artificial Intelligence in Communication Disorders, andreas.kist@fau.de