AphaSpeak: A Real-Time Multimodal System to Enhance Expressive Communication for Individuals With Aphasia

Lin, Xuan Yu (School: National Hualien Senior High School)
Chuang, Chia-Wei (School: National Hualien Senior High School)

Individuals with non-fluent aphasia, such as Broca's aphasia, face profound communication barriers due to impaired verbal expression. Traditional assistive tools often rely on static image databases, which lack contextual flexibility and involve high cognitive loads. This research develops AphaSpeak, an innovative AI-driven multimodal communication system integrated into an iOS application. The system processes environmental imagery, speech, lip movements, gestures, and emotions, utilizing Large Language Models (LLMs) to reconstruct fragmented inputs into fluent narratives. To address the scarcity of Chinese aphasic corpora, this study developed AphasiaSim-LLM, a novel data generation method using BERT-based attention masking and QLoRA fine-tuning, achieving high realism validated by Naive Bayes classifiers. A quantitative evaluation framework identified Gemini 2.5 Flash as the optimal model for semantic restoration. A major engineering breakthrough of this study is the implementation of a parallelized streaming architecture to enable synchronous multimodal computation. By transitioning from serial processing to a streaming pipeline, combined with FFmpeg for rapid video decoding and ORB-based keyframe extraction, AphaSpeak successfully eliminated computational bottlenecks. This optimization significantly reduced total system latency to under 4 seconds, ensuring the responsiveness required for natural social interaction. Clinical validation using the Aphasia Bank dataset confirms that this high-speed, personalized system effectively empowers aphasia patients with fluid and expressive communication capabilities.

Awards Won: