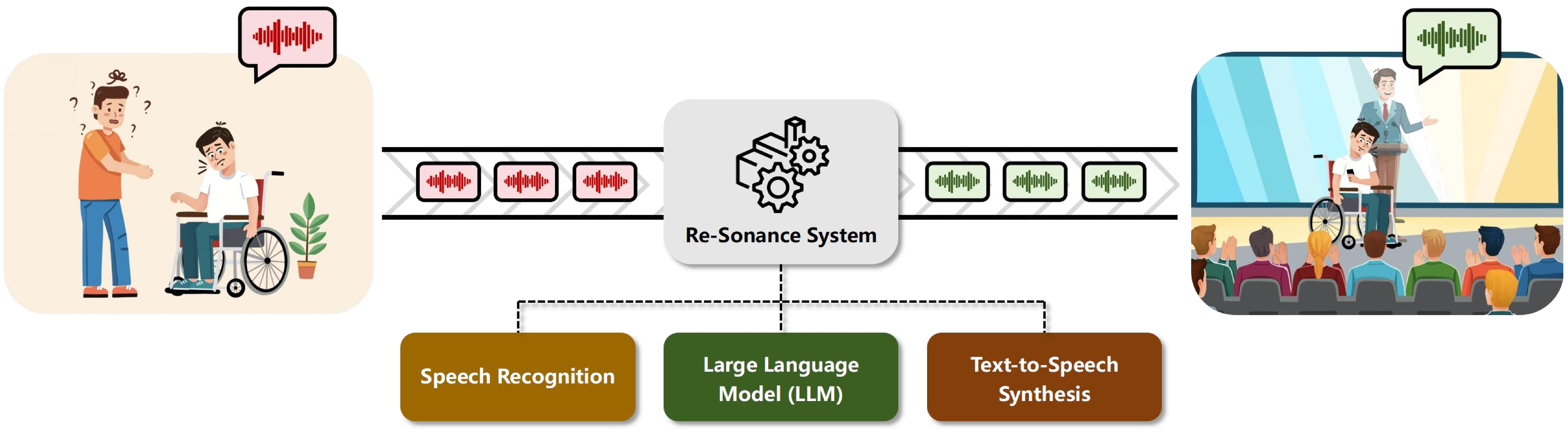
# Re-Sonance: A Dysarthric Asynchronous Real-Time Speech Conversion System Based on a Three-Stage Cascaded ASR-LLM-TTS Architecture

Yuxuan Wu, Yifan Xu, Junkun Wang, Jiayong Jiang, Xin Zhao, Zhaojie Luo
[1] State Key Laboratory of Digital Medical Engineering, Southeast University, Nanjing, China
[2] School of Biological Science & Medical Engineering, Southeast University, Nanjing, China




## Introduction

Dysarthria — a neuromotor speech disorder caused by congenital or acquired conditions — severely impairs articulation and real-time communication, reducing quality of life and social participation.

While modern Augmentative and Alternative Communication (AAC) systems have improved expression, most rely on speech replacement, which often disrupts natural communication flow and lacks real-time responsiveness.

Such unnatural interaction can diminish users' self-identity, social confidence, and motivation to engage with assistive tools. Moreover, system latency and operational complexity make them unsuitable for dynamic contexts such as conferences or presentations.

To overcome these challenges, we introduce Re-Sonance, a real-time, speech-driven AAC system enabling natural participation through minimally intrusive interaction. It combines ASR for dysarthric speech recognition, LLMs for contextual refinement, and TTS for natural, intelligible voice reconstruction.

Our work bridges critical gaps in AAC technology, advancing accessible, speech-driven communication and promoting linguistic equity for individuals with dysarthria.

## Method

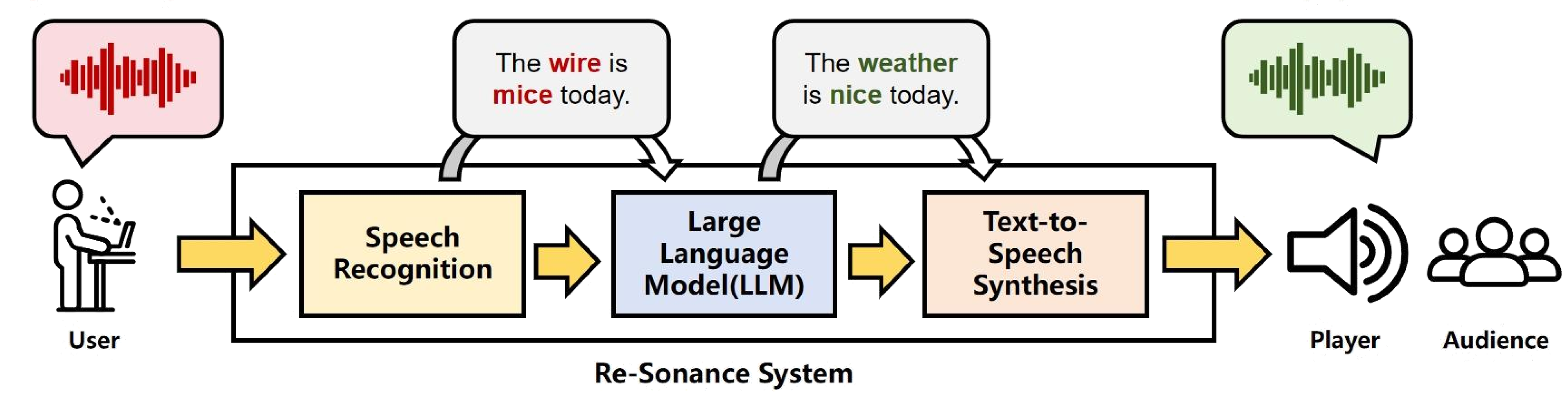



*Fig. 1. Technical framework and application scenarios of Re-Sonance*

**System Overview**

Re-Sonance converts dysarthric speech into clear, natural output through a three-stage cascaded architecture integrating:

1. ASR: Recognizes dysarthric speech and generates partial text streams.
2. LLM: Refines and corrects ASR output via prompt-based contextual inference.
3. TTS: Synthesizes natural, intelligible speech with multilingual and personalized voice support.

**Asynchronous Real-Time Design**

Implements a non-blocking pipeline where ASR, LLM, and TTS operate in parallel streams.
Partial ASR results are immediately processed by the LLM and TTS, achieving low latency (RTF ≈ 0.82).
Enables smooth, near-instantaneous feedback during real-time communication.

**Interface & Deployment**

Optimized for mobile and desktop platforms with simple input/output control.
Designed for professional contexts such as meetings and presentations.
Public demo available: demo-resonance.hai-lab.cn

## Result

### Subjective Evaluation

20 native Mandarin speakers rated intelligibility, naturalness, and semantic relevance of speech samples. Re-Sonance markedly enhances clarity and perceived quality for mild-to-moderate users.

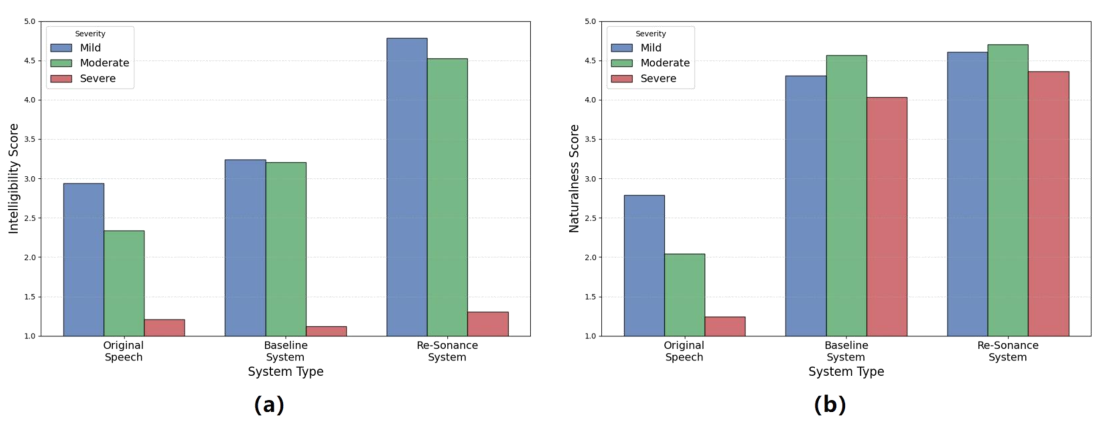



*Fig. 3. Distribution of (a) Intelligibility and (b) Naturalness scores across system types*

### Objective Evaluation

Dataset: Chinese Dysarthric Speech Dataset (CDSD).
Metrics: Word Error Rate (WER), Match Error Rate (MER), Word Information Lost (WIL).

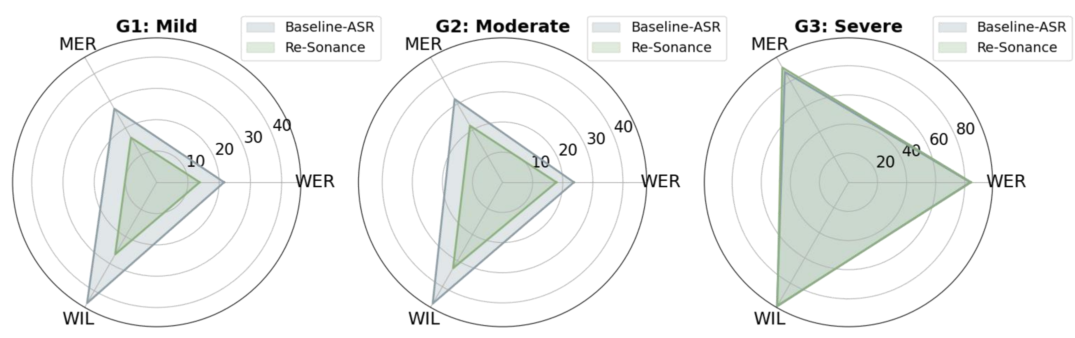



*Fig. 4. Comparison of Baseline-ASR and Re-Sonance for Different Severity Levels*

### Latency

Real-Time Factor (RTF): 0.82 → system operates faster than real time, enabling smooth conversational flow.
Confirms the feasibility of asynchronous streaming for real-world AAC applications.

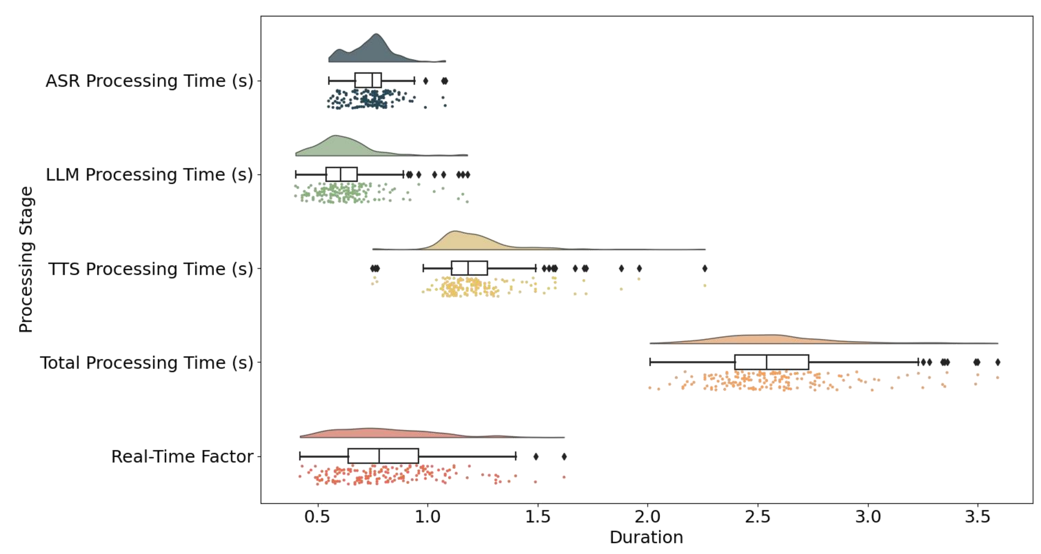


*Fig. 5. Distribution of Processing Times Across Different Stages*

## Conclusion

Integrating large language models (LLMs) into traditional ASR-TTS pipelines significantly enhances speech intelligibility, naturalness, and semantic coherence for individuals with mild to moderate dysarthria. The LLM component effectively corrects ASR errors and infers missing context without model fine-tuning, demonstrating strong adaptability through prompt optimization. However, performance for severe dysarthria remains limited, and the system's evaluation in Mandarin only restricts cross-linguistic generalization. Although latency is minimized through asynchronous streaming, slight inference delays persist. Future work will focus on personalized ASR training, multilingual validation, and longitudinal user studies. Overall, Re-Sonance advances LLM-based assistive communication, promoting natural, real-time, and inclusive speech interaction for people with dysarthria.