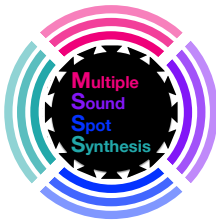


Speech Masking System

Based on Spatially Separated Multiple TTS Maskers With A Compact Circular Loudspeaker Array

Takuma Okamoto

National Institute of Information and Communications Technology, Japan



Related our work



Interspeech 2025 Show & Tell



Multiple sound spot synthesis project HP



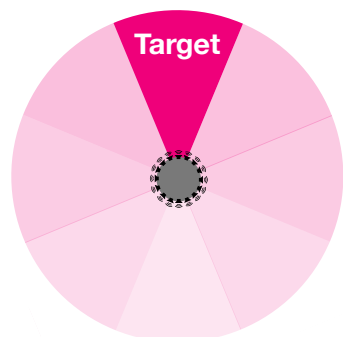
Demo at EXPO 2025

1. Introduction

- Speech privacy and sound masking
 - Localized sound spot synthesis
 - > target sound can be heard even slightly in non-target direction
 - Speech masking system with time-reversal signals of target speech
 - > Meaningless sounds are difficult to hear
 - > Target speech can be sometimes heard

Proposed demo system

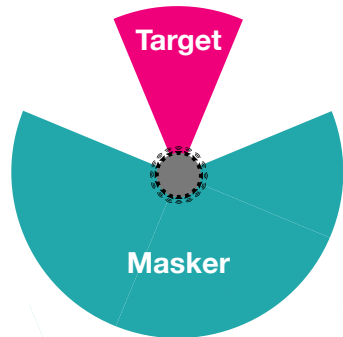
- Speech masking system with text-to-speech (TTS) and multiple sound spot synthesis
- Compact demo system carried out with a backpack



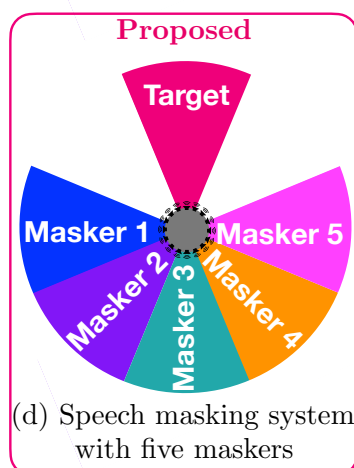
(a) Localized sound spot synthesis



(b) Multiple sound spot synthesis



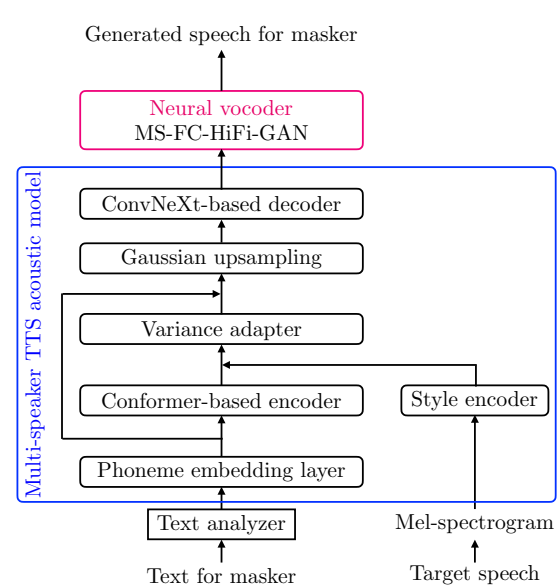
(c) Speech masking system with single masker



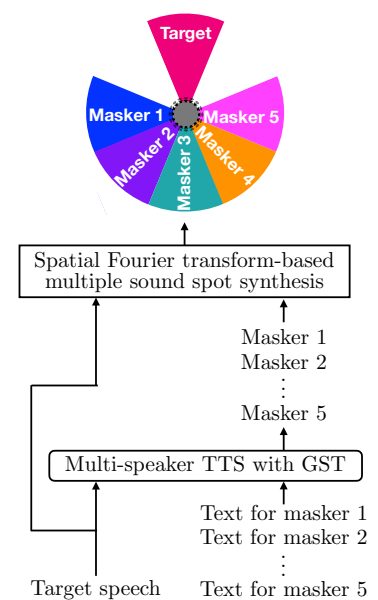
(d) Speech masking system with five maskers

2. Proposed speech masking system

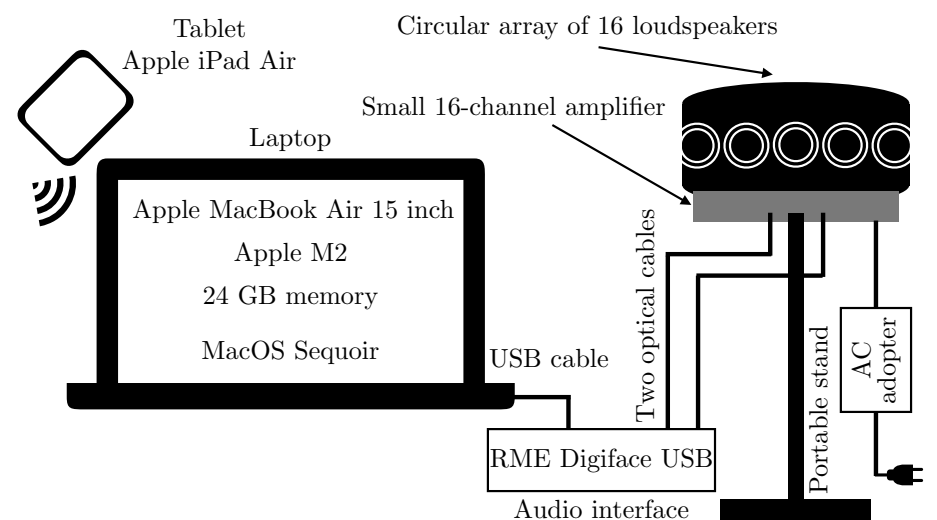
- Proposed spatially separated multiple TTS-based maskers
 - Maskers with same voice quality as target speech are generated with multi-speaker TTS
 - Target speech and generated five TTS-based maskers are synthesized with multiple sound spot synthesis



(a) Multi-speaker TTS with global style tokens



(b) Proposed speech masking system



Demonstrations with an implemented system

- Localized sound spot synthesis
- Multiple sound spot synthesis
- Speech masking system with single masker
- Speech masking system with five TTS maskers

Acknowledgement: This study was partly supported by JSPS KAKENHI Grant Number JP23K11177.

