AI-Powered
Target Speaker Extraction

Overview:

Meeami's Target Speaker Extraction (TSE) technology isolates a specific speaker's voice from a complex audio scene — even in the presence of other voices, background noise, or overlapping speech. Using a single clean enrollment clip from the target speaker, TSE reconstructs their speech with high clarity while suppressing all other sound sources.

Target Use Cases

- **Call Centers & Voice Analytics**
  Identify and isolate agent or customer speech for quality assurance, transcription, and compliance.

- **Online Meetings & Collaboration**
  Suppress background participants and highlight the enrolled speaker — useful in noisy shared environments.

- **Smart Speakers & Voice Assistants**
  Focus on the authorized user's voice even in multi-user households or noisy conditions.

- **Security & Forensics**
  Extract suspect or witness speech from surveillance audio.

Technical Specifications

Feature	Specification
Latency	~10 ms
Model Size	22 MB total
- Speaker Embedding Module	10 MB
- Extraction Module	12 MB
Compute	~600M MACs
Input Type	Single-channel audio
Enrollment Required	Yes (1-time, clean sample)
Output	Clean voice of enrolled speaker

How it works

he TSE system consists of two modules:

1. **Speaker Embedding Model**
   - Takes a clean 3-5 sec enrollment sample
   - Generates a compact voiceprint representation
2. **Speaker Extraction Model**
   - Uses the voiceprint to filter incoming audio
   - Works in real time using frame buffers

Integration

- **Meeami SDK (Under Optimization Phase)**
  - Full integration with Meeami's CleanAudio stack (AEC, NS, Speaker ID)
- **Deployment Targets:**
  - Call center software and telephony stacks
  - Embedded platforms (smart speakers, custom silicon)
  - Conferencing platforms and SDKs

Outcomes

- **Clear Voice Extraction in Overlapping Speech Scenarios**
  Even in crowded environments, extract the intended voice with clarity.

- **Improved ASR/Analytics**
  Feed clean voice into downstream pipelines like speech-to-text, emotion recognition, or intent detection.

- **Low-Latency Processing for Real-Time Applications**
  Ensures that call quality and voice UX are never compromised.

Transformational Impact

Without TSE	With TSE
Multiple voices confuse ASR engines	Only enrolled speaker's voice is preserved
Difficult to isolate target speaker during analysis	Clean voice output enables accurate analytics
Noisy audio in shared spaces	Voice isolation enhances clarity and privacy
Microsoft Voice Isolation requires 30s training	Meeami's TSE works with just one clean enrollment

Talk to Sales or Request a Demo

Meeami's TSE is ready for real-time deployments and PoC integrations
sales@meeamitech.com