

Two-Stage Speaker Anonymization with Emotion-Conditioned Resynthesis and Mild McAdams Post-Processing

A COSYVOICE2 SYSTEM FOR THE VOICEPRIVACY 2026 CHALLENGE

TEAM-CHAMELEOS

Track 1: English anonymization, VoicePrivacy 2026 Challenge · in conjunction with SPSC 2026, Sydney

Jiarui Xu

ShanghaiTech University

xujr2025@shanghaitech.edu.cn

Zhiyu Wan

ShanghaiTech University

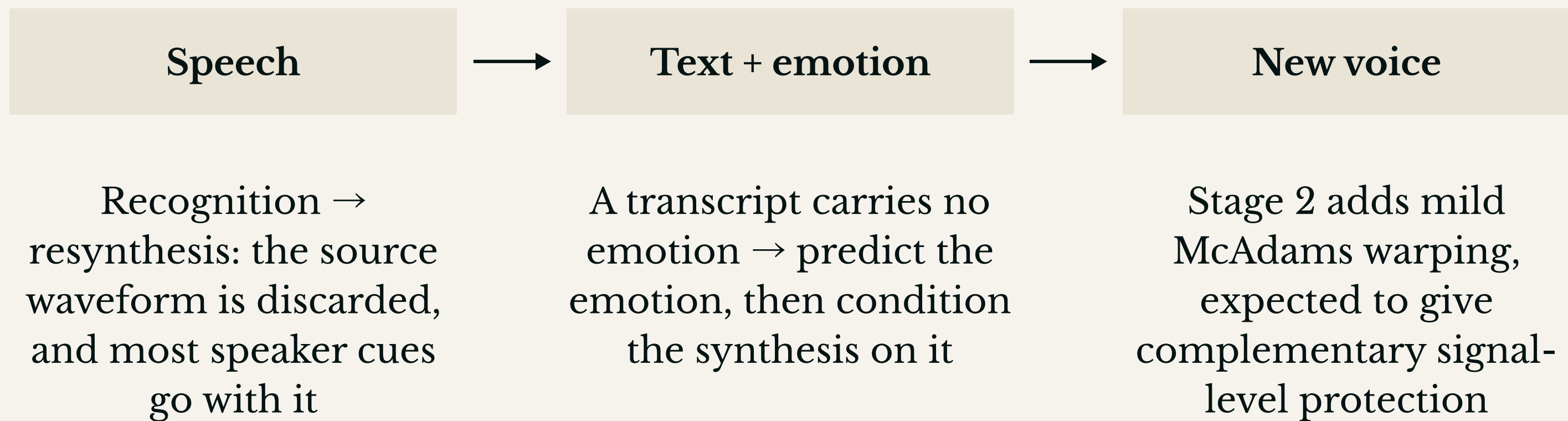
wanzhy@shanghaitech.edu.cn

Sihan Xie

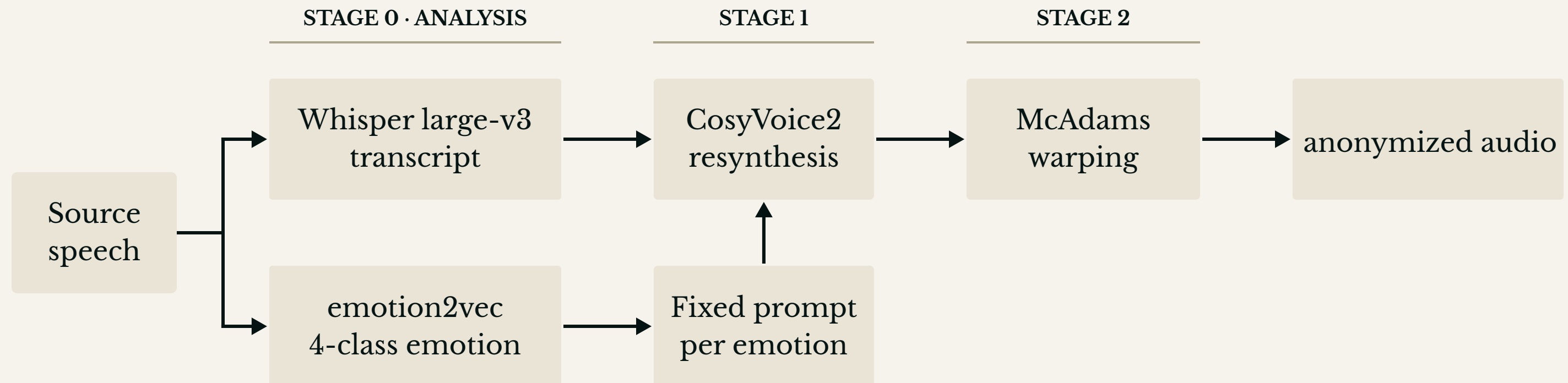
ShanghaiTech University

xiesh2024@shanghaitech.edu.cn

Core idea



Pipeline



- Only the transcript and the emotion class reach the synthesizer; no source audio does.
- Utterance-level: the same procedure per utterance; no ground-truth transcripts, emotion labels or speaker labels

Stage 1 design choices

Class	emotion2vec labels	CosyVoice2 path
ang	angry, disgusted	instruct2 Angry-style instruction
hap	happy	zero-shot Prompt audio + transcript
neu	neutral, other, surprised, unknown	zero-shot
sad	sad, fearful	zero-shot

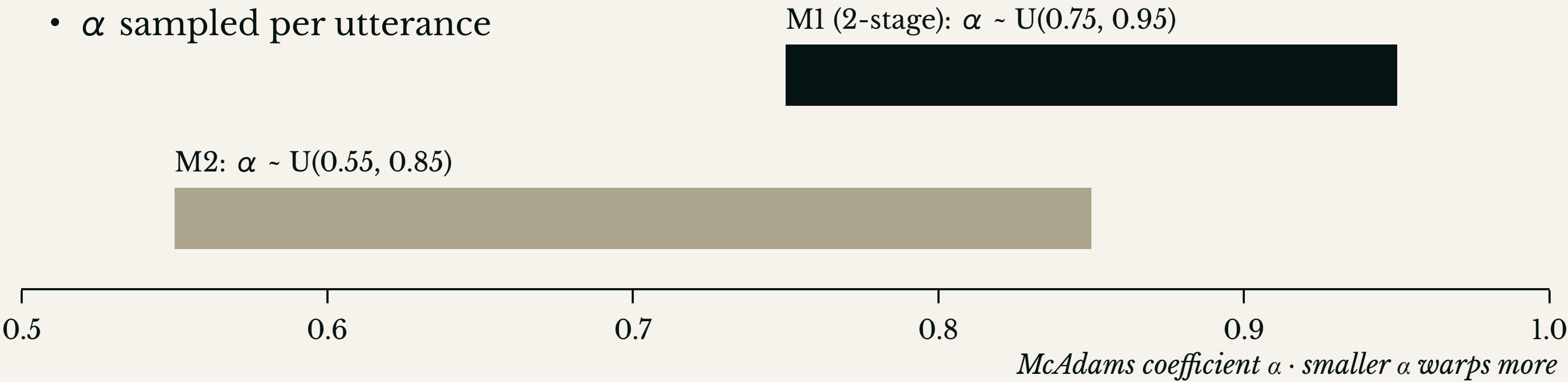
Reference audio

Class	Reference clip and transcript	Path	Preparation
ang	1091_IEO_ANG_HI “It's eleven o'clock.”	instruct2	No processing
hap	1074_IEO_HAP_MD “It's eleven o'clock.”	zero-shot	Normalize to -26 dBFS
neu	1077_IOM_NEU_XX “I'm on my way to the meeting.”	zero-shot	F0 range $\times 0.7$ Compression 1.5:1 Level -30 dBFS
sad	1068_ITS_SAD_XX “I think I've seen this before.”	zero-shot	Level -42 dBFS Compression 2:1

Selection: regress reference success rate on loudness, dynamics, F0 range, speaking rate and reference SER predictions.

Stage 2: mild McAdams

- McAdams transformation of the LPC spectral envelope: shifts formant structure, alters vocal-tract cues
- 20 LPC coefficients
- 20 ms window
- 10 ms shift
- α sampled per utterance



Results

System	UAR	WER	EER_base	EER_semi
Original speech	69.15 / 71.06	1.80 / 1.84	–	–
2-stage system	66.09 / 65.97	4.55 / 4.29	44.57 / 43.75	35.90 / 32.40
Stage-1 only	67.73 / 68.14	4.14 / 3.96	44.82 / 43.55	35.54 / 32.58
Stage-1 only with fine-tuned CREMA-D prompts	72.13 / 73.78	4.18 / 3.99	43.62 / 43.53	35.75 / 31.31

McAdams ablation

Change in percentage points: 2-stage system minus Stage-1 only

Metric	Dev change	Test change
UAR	-1.64	-2.17
WER	+0.41	+0.33
EER_base	-0.25	+0.20
EER_semi	+0.36	-0.18

Utility declines on both splits
Privacy changes are small and inconsistent

Findings and limitations

Limitations

- Test WER: 1.84% original, 4.29% 2-stage, 3.99% v3
- Test EER_semi below 40%: 32.40% 2-stage, 31.31% v3

Next

- Validate prompt selection across more reference recordings
- Improve happy / neutral recall and adapted-attacker privacy

Key observations

Stage 1 provides most of the privacy protection in the 2-stage system

McAdams reduces utility without consistent privacy gains

v3 improves UAR, but test EER_semi falls to 31.31%

Thank you

CODE AVAILABLE AT:

github.com/AnDuinwryn/vpc2026-cosyemotion-mcadams