Seq2Seq TTS中编码器结构的探索

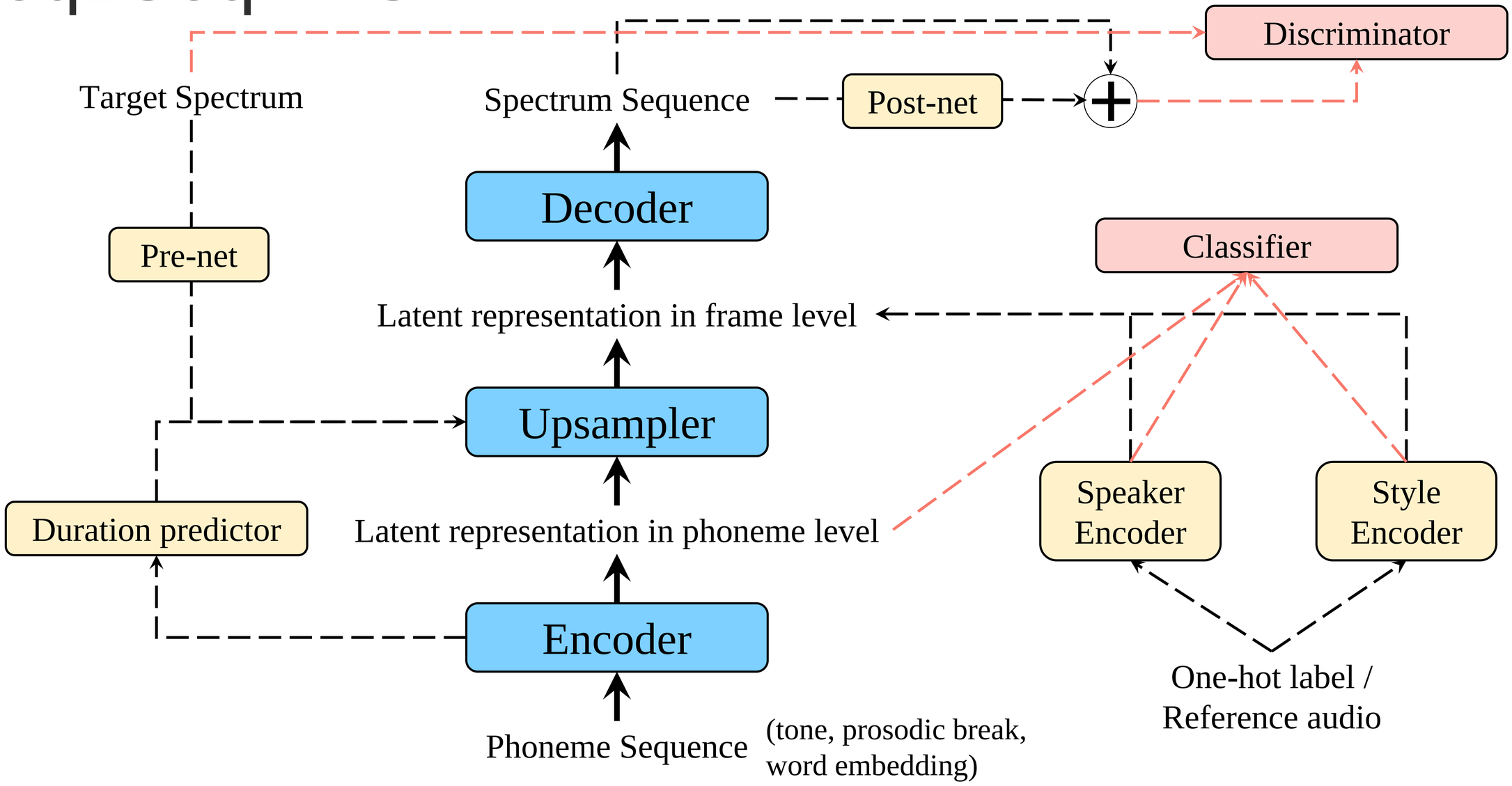
栾剑

Outline

- Seq2Seq TTS
- Main stream encoders
- Proposed aggression structure
- TTS applications in Xiaomi *Update*

luanjian@xiaomi.com

Seq2Seq TTS



luanjian@xiaomi.com

CBHG @ Tacotron

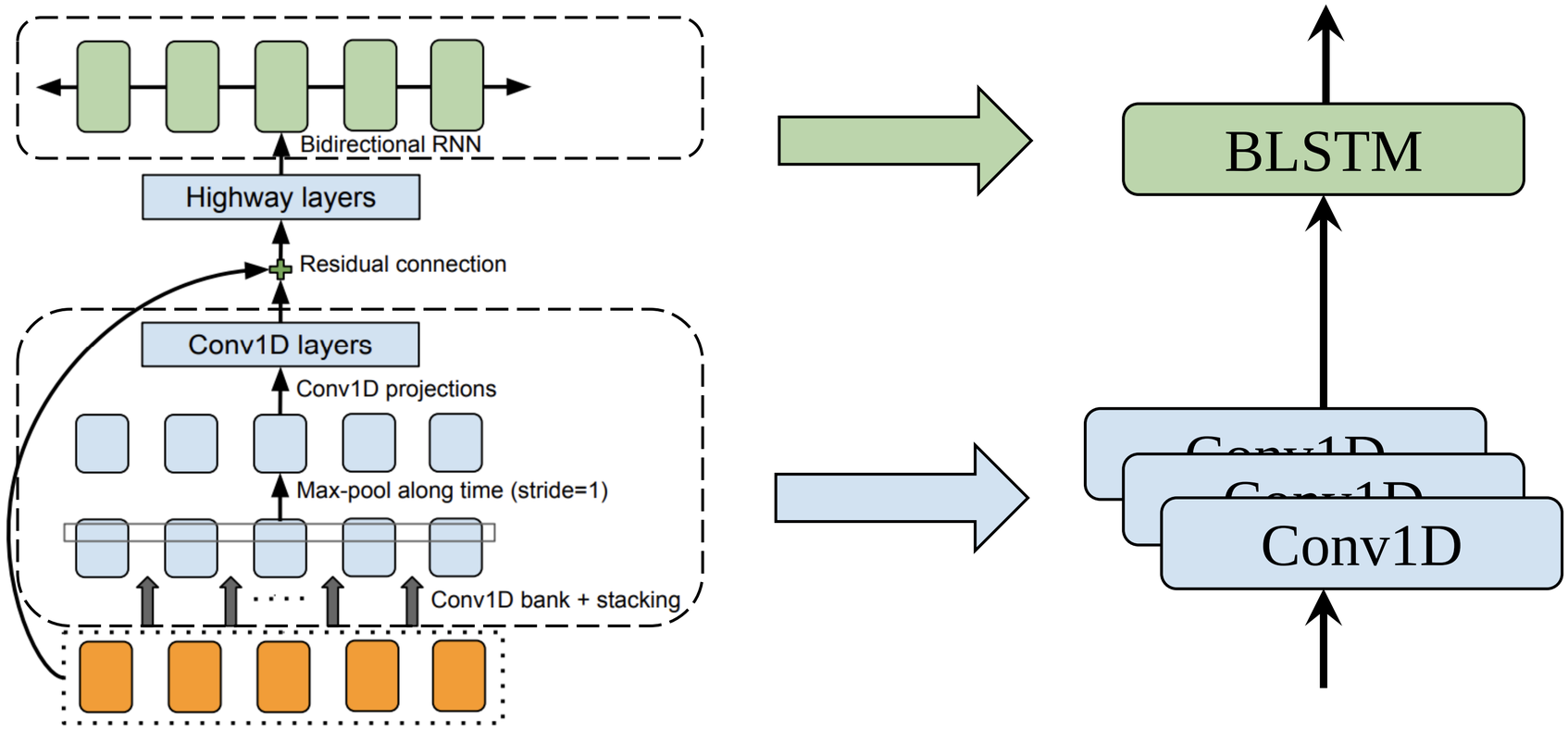
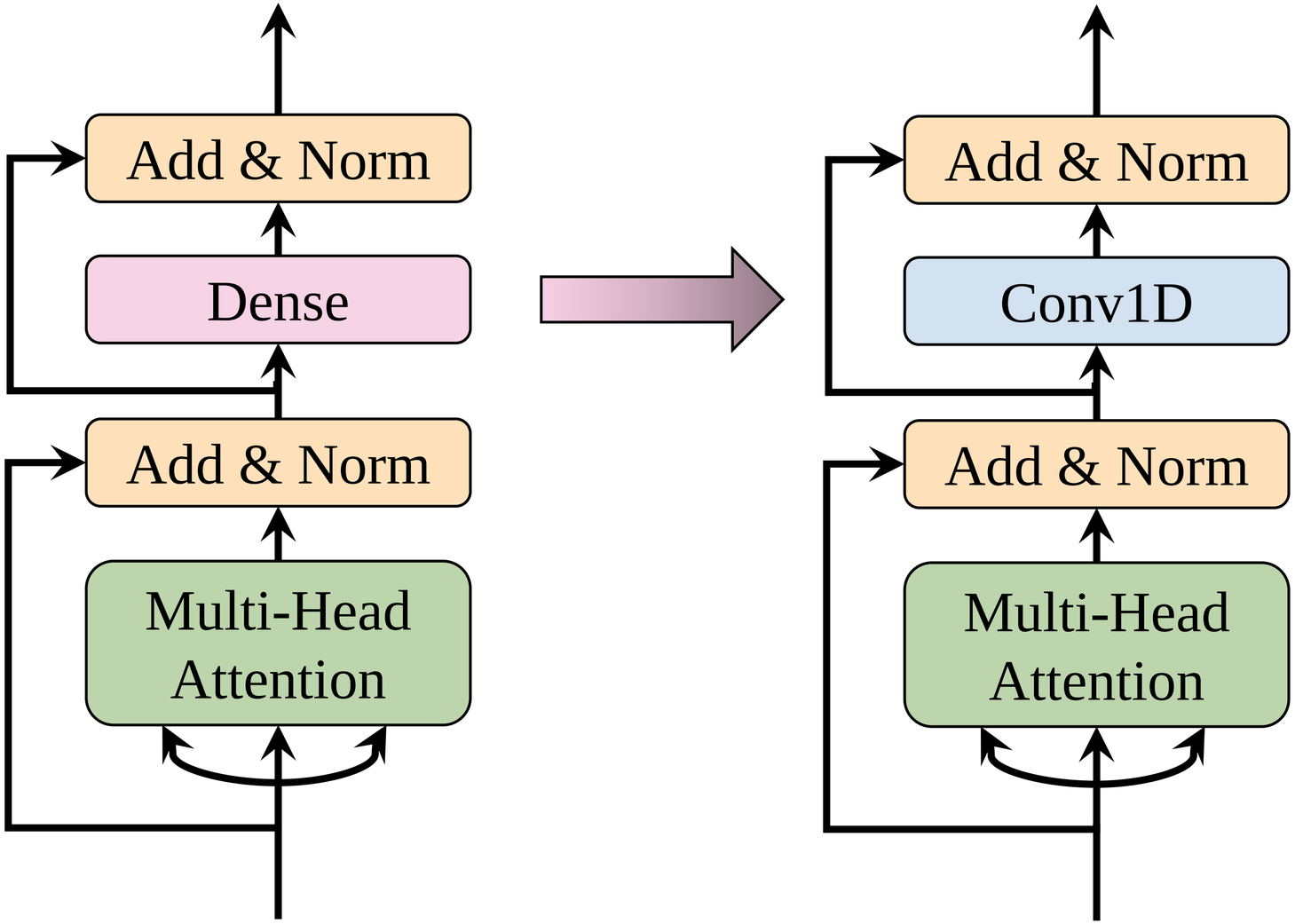


Figure 2: The CBHG (1-D convolution bank + highway network + bidirectional GRU) module adapted from Lee et al. (2016).

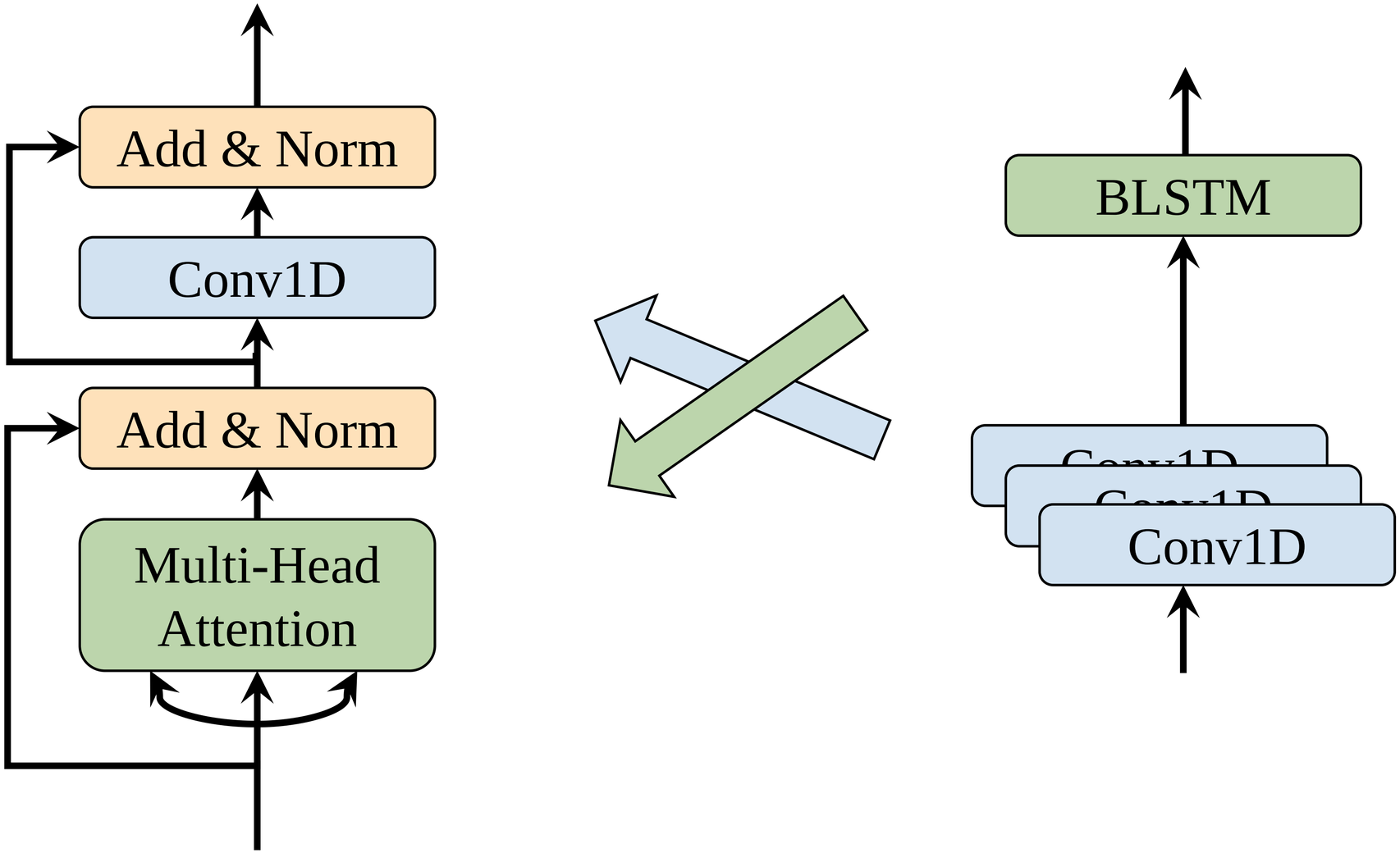
luanjian@xiaomi.com

SAN @ Transformer



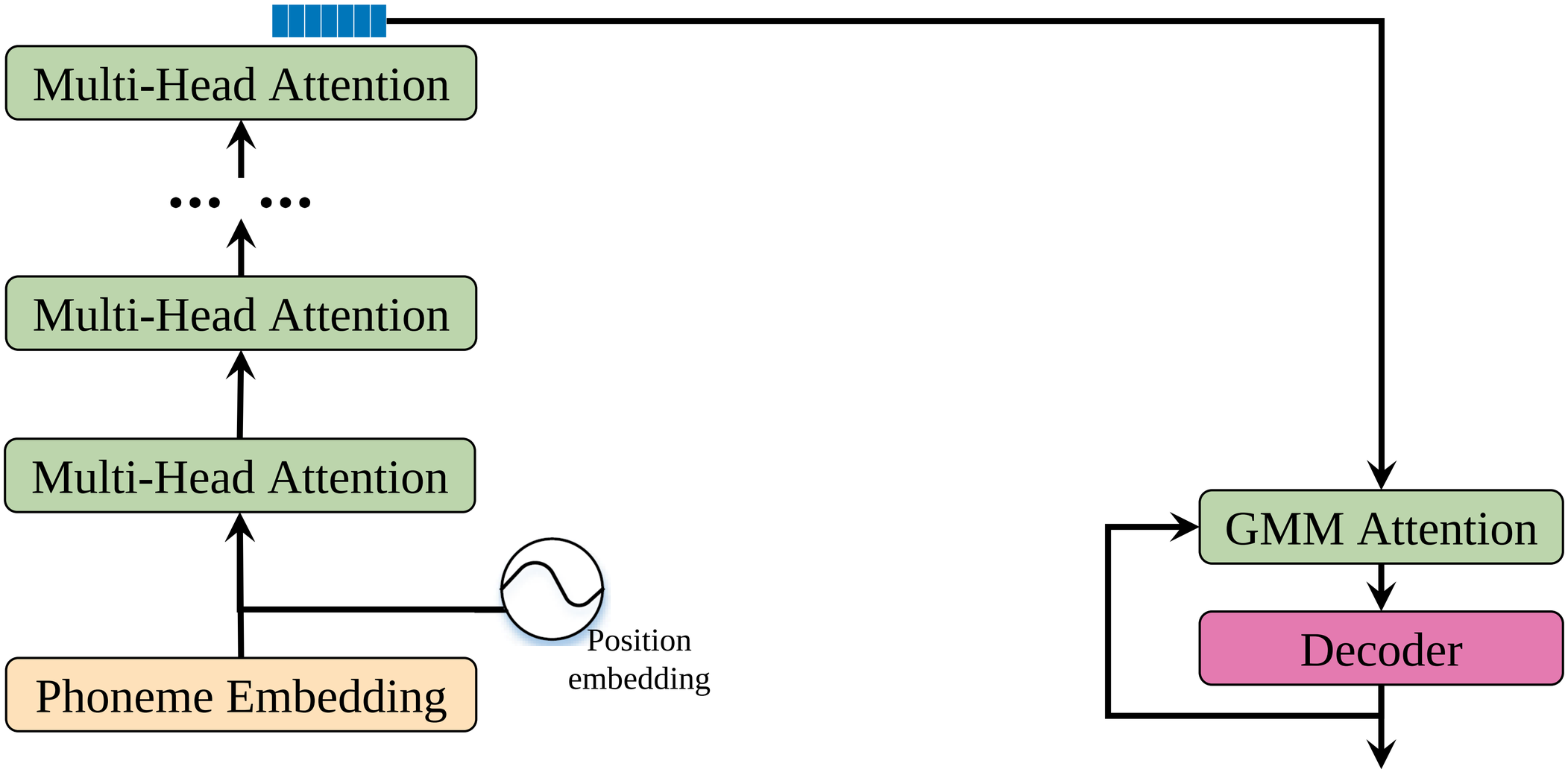
luanjian@xiaomi.com

FFT-Block @ FastSpeech



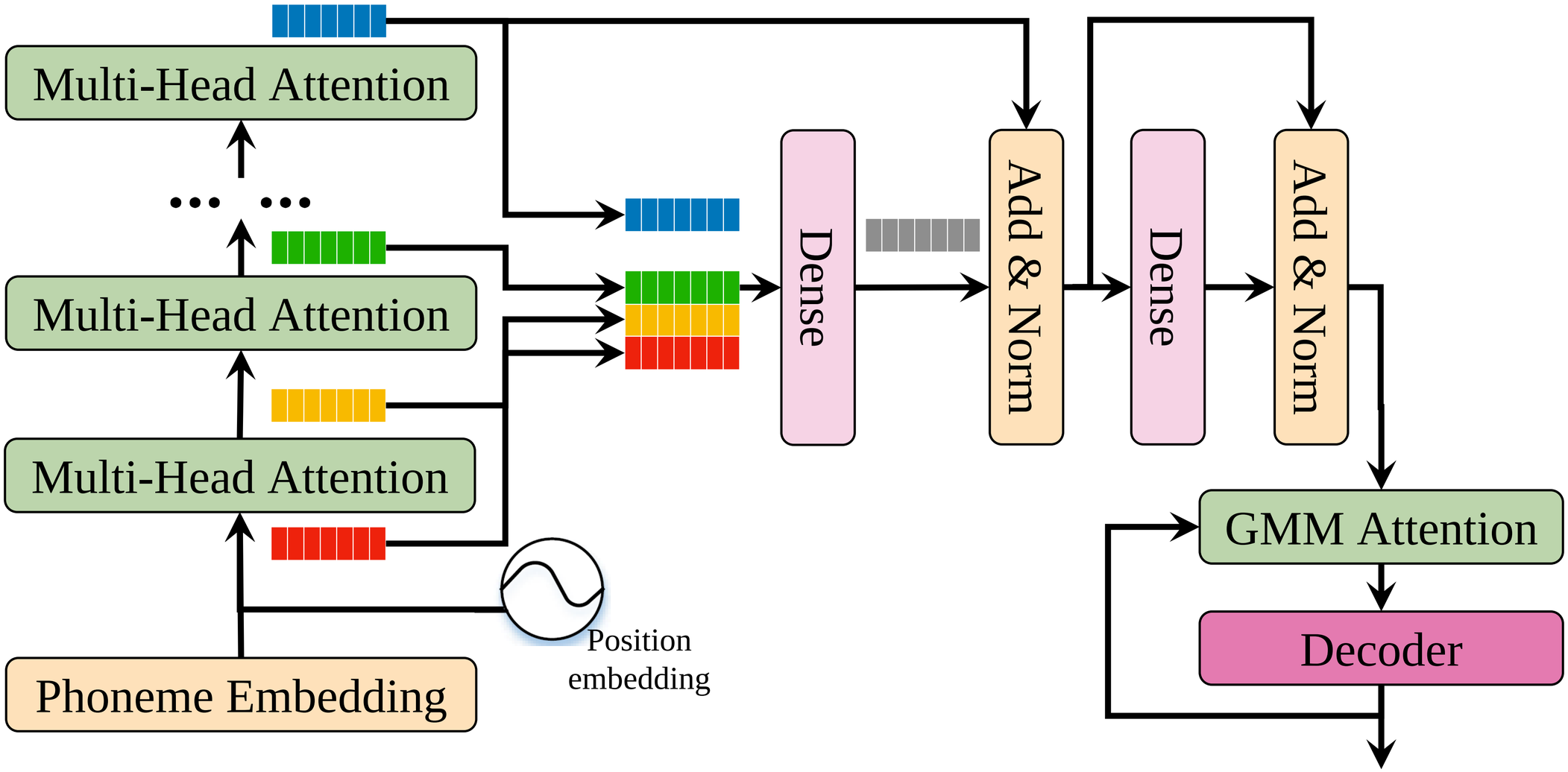
luanjian@xiaomi.com

Baseline



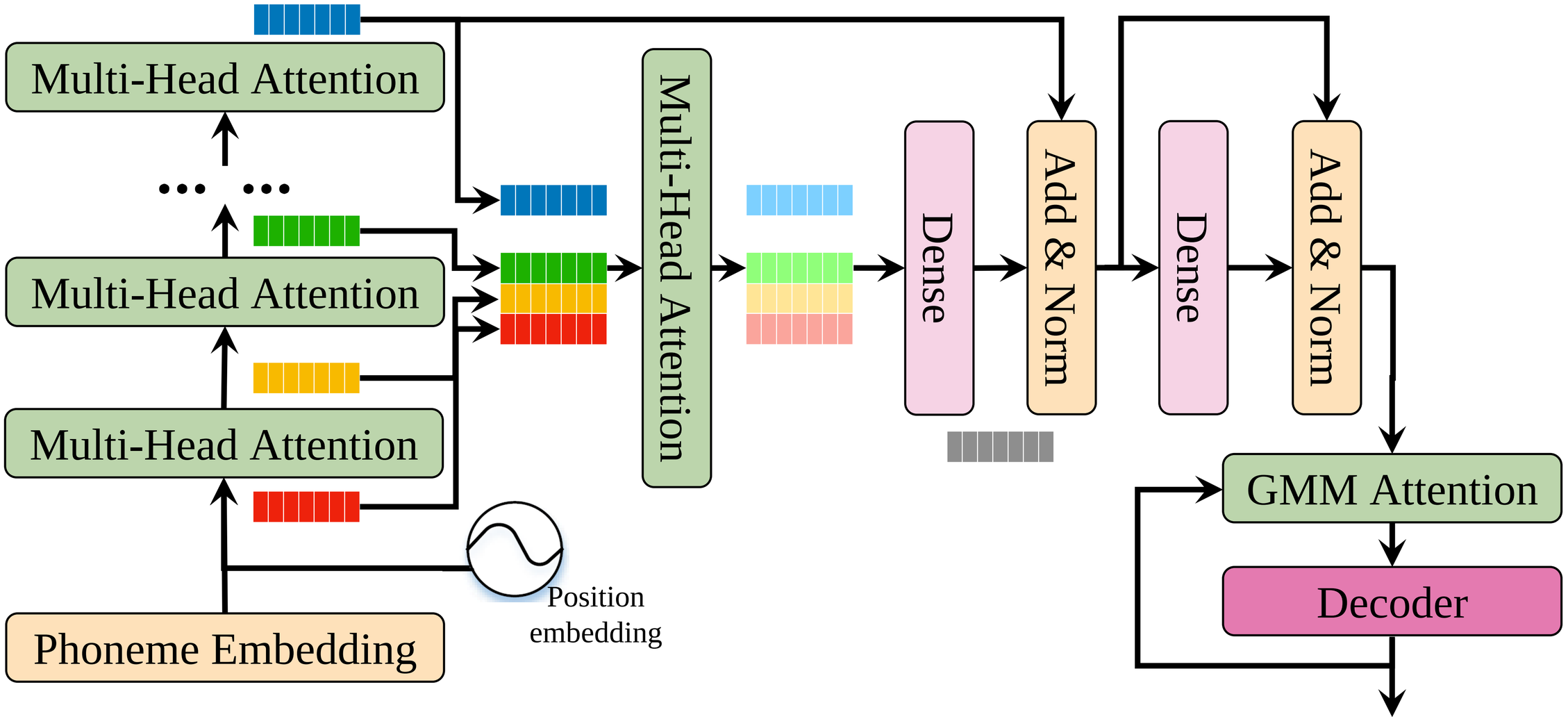
luanjian@xiaomi.com

Direct Aggregation



luanjian@xiaomi.com

Self-Attention-Based Aggregation



luanjian@xiaomi.com

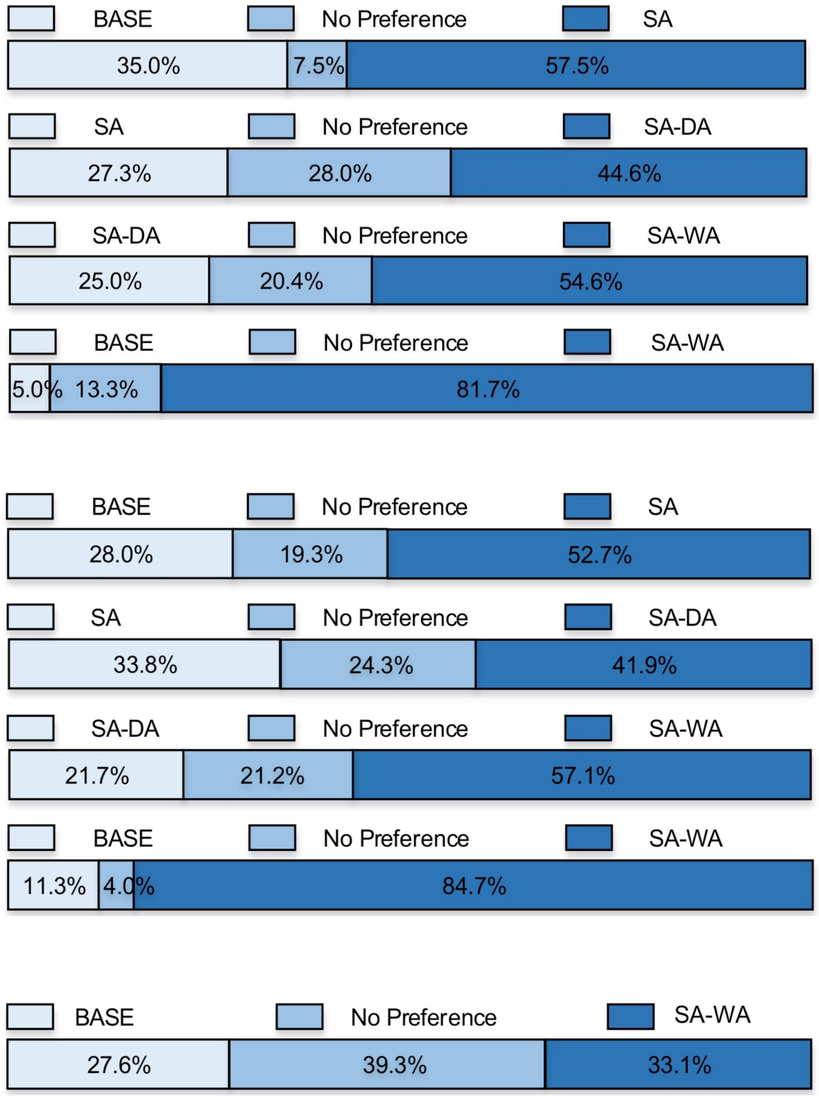
Experiments

数据：脱口秀(TS)、语音助手(VA)、朗读风格(DB1)

系统	Encoder
Base	CBHG
SA	SAN
SA-DA	SAN + Direct Aggression
SA-WA	SAN + Self-Attention-based Aggression

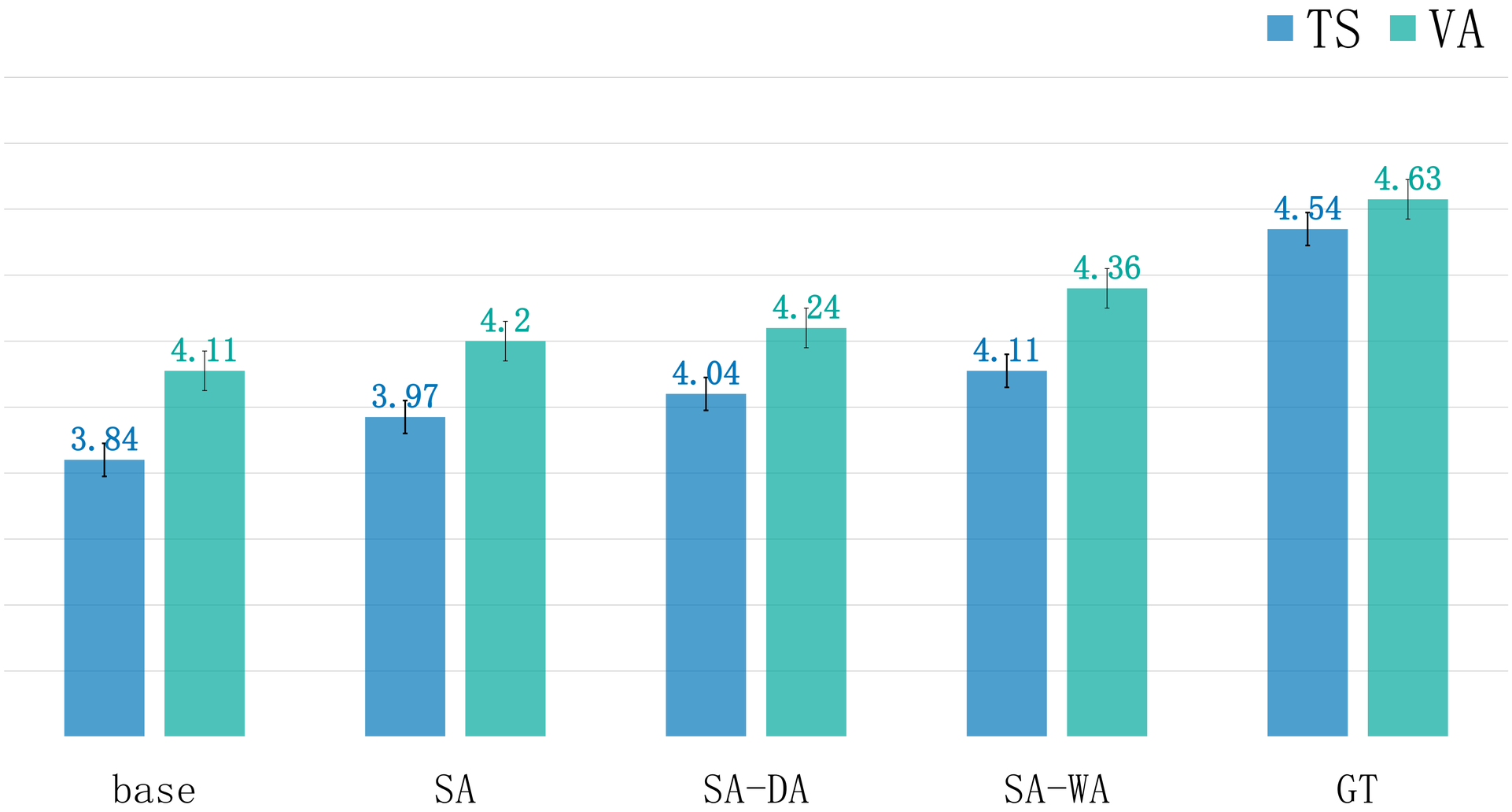
MCD

语料库	BASE	SA	SA-DA	SA-WA
TS	8.01	7.48	7.42	**7.32**
VA	7.60	7.37	7.32	**7.23**



luanjian@xiaomi.com

MOS



luanjian@xiaomi.com

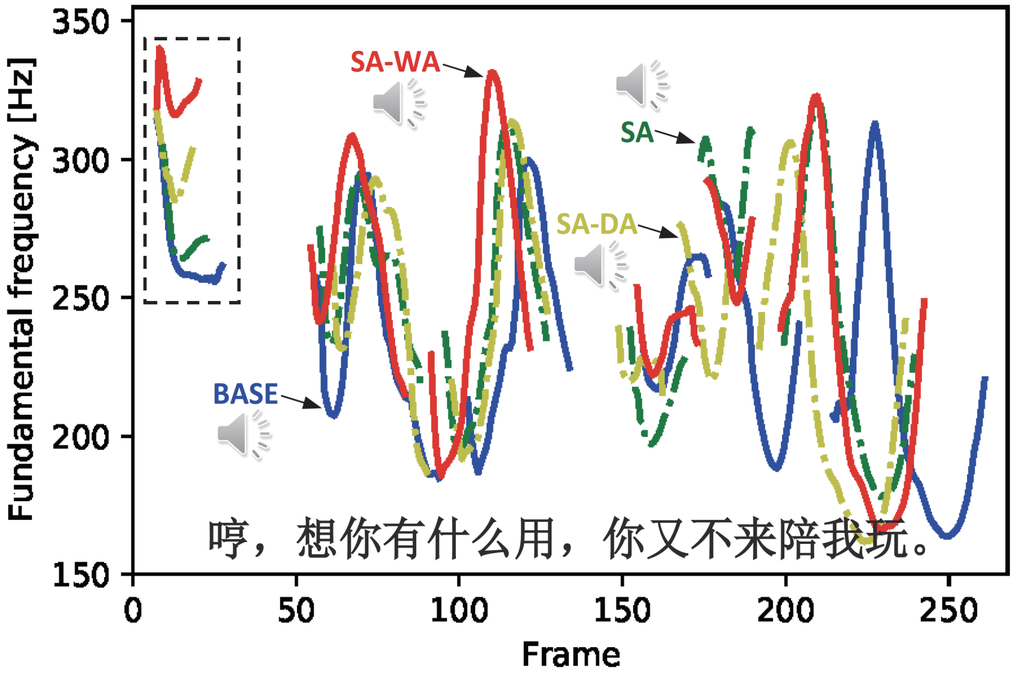
Experiments

- **韵律相关性：**

	BASE	SA	SA-DA	SA-WA
E	0.755	0.776	0.781	**0.799**
Dur.	0.617	0.638	0.641	**0.654**
F0	0.42	0.426	0.437	**0.501**

- **韵律多样性：**

	BASE	SA	SA-DA	SA-WA	GT
E	0.238	0.277	0.285	**0.304**	0.321
Dur.	33.374	34.337	34.955	**37.003**	41.866
F0	32.362	33.405	35.161	**35.766**	36.824



* 一句话内每个音素内的相对能量(E)、时长(Dur.)和基频(F0)

luanjian@xiaomi.com

声音体验是小爱同学的一大特色

官方音色：

声音商店：

luanjian@xiaomi.com

自研TTS

覆盖所有手机、音箱、电视及大部分MIoT设备

日请求量 2.2亿+

月活 0.9亿+

luanjian@xiaomi.com

探索更多能力

歌唱

编曲

朗诵

luanjian@xiaomi.com

#雷军和小爱同学的父亲节互动

luanjian@xiaomi.com

Thanks

luanjian@xiaomi.com