HWB-plus: Lightweight Speech Bandwidth Extension with Separate Modeling for Consonants and Vowels

Xin Liu | Xueliang Zhang

College of Computer Science, Inner Mongolia University, China | xin.liu@mail.imu.edu.cn | cszxl@imu.edu.cn

Core Contributions

01 Source-aware envelope shaping

Drawing on STRAIGHT, HWB-plus separates periodic vowels from aperiodic consonants. Each WGMM shapes its own source-aligned base: HWR-enriched speech or white noise.

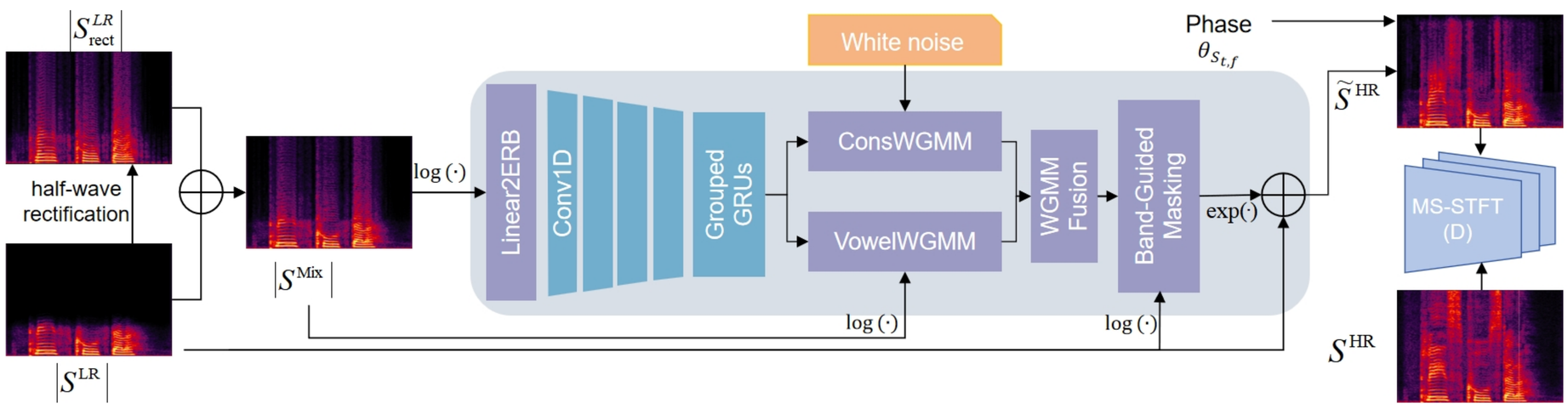
02 Mel-scale spectral prior

VowelWGMM initializes Gaussian centers and bandwidths from mel-filter geometry, focusing capacity on perceptually meaningful regions.

03 Same deployment budget

DualWGMM replaces the single WGMM while retaining 194K parameters. Computation changes only from 12.33 to 12.39M MACs/s.

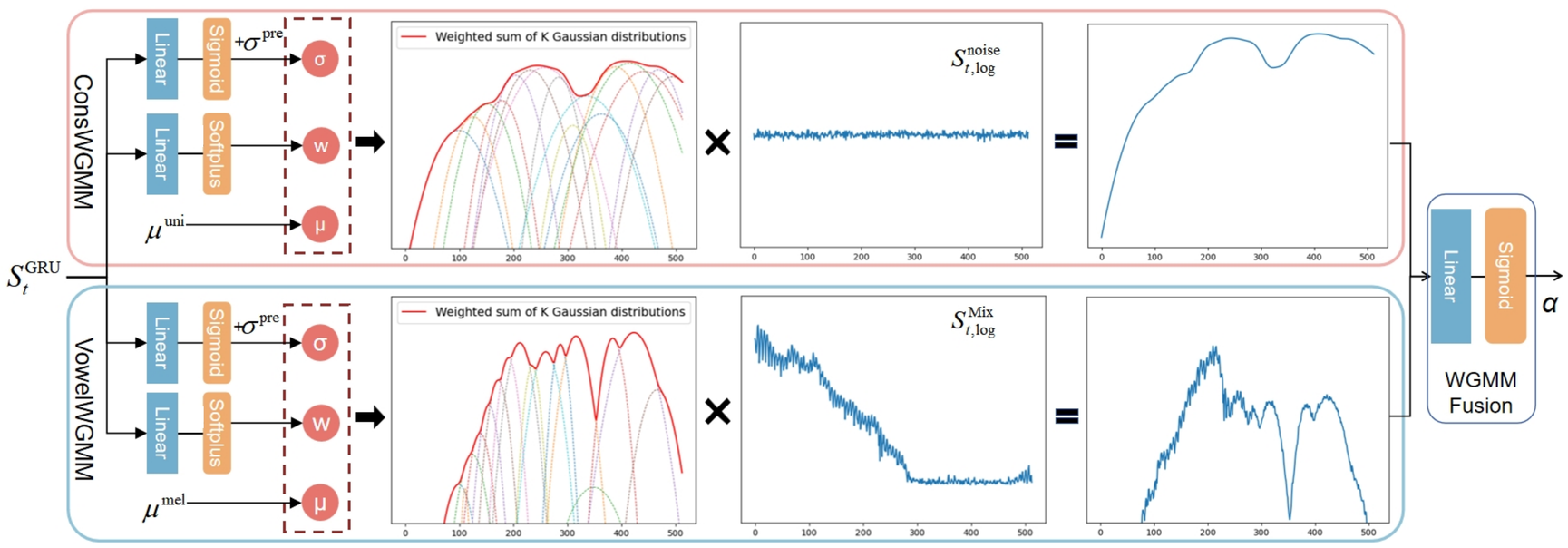
End-to-End System



The HWB encoder is retained. DualWGMM replaces the single spectral generator before band-guided masking.

DualWGMM: Source-Specific Modeling and Learned Fusion

The two branches use distinct base signals and parameter initialization, then a learned frame-level fusion combines their HF estimates.



$$S_{\log(pred)} = \alpha \cdot S_{\log(vowel)} + (1-\alpha) \cdot S_{\log(cons)}$$

Experimental Setting VCTK, 110 speakers: 90 train / 10 dev / 10 test. LR: Butterworth low-pass, 2-3 kHz cutoff. HR target: 22.05 kHz.

Results at the Same Lightweight Scale

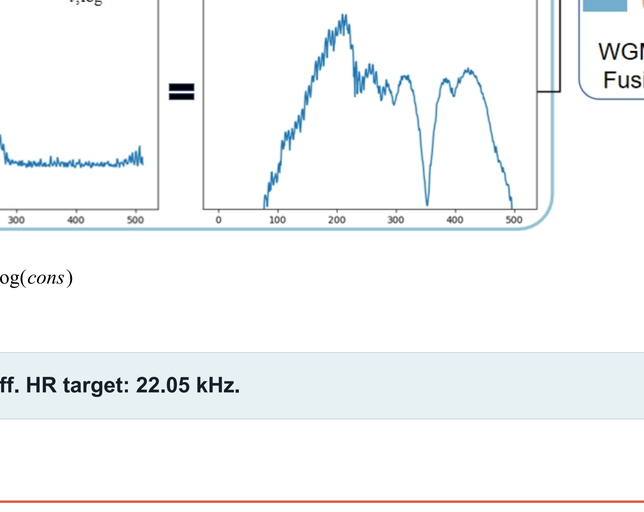
Method	LSD ↓	P.808 ↑	PESQ ↑	VISQOL ↑	NISQA ↑	Para. K	MACs M/s
Sinc	3.19	2.99	3.30	2.84	3.15	-	-
BAE	0.84	3.34	3.78	3.72	3.76	2827	246
BAE-Lite	0.86	3.25	3.37	3.79	3.70	569	26.23
HWB	1.11	3.31	3.35	3.52	3.69	194	12.33
HWB-plus	**0.94**	**3.55**	**3.83**	**3.96**	**3.93**	**194**	**12.39**

Ablation: Each Design Component Contributes

Method	LSD ↓	P.808 ↑	PESQ ↑
HWB-plus	**0.94**	**3.55**	**3.83**
without DualWGMM	0.94	3.36	3.64
without mel initialization	1.07	3.49	3.62
without log domain	0.97	3.31	3.54

Removing DualWGMM keeps LSD at 0.94 but reduces listening-oriented scores. Mel-scale initialization and log-domain processing also improve performance.

Spectral Evidence



DEMO WEBSITE

Source-aligned WGMM envelopes model periodic vowel and aperiodic consonant components separately, improving perceptual quality without increasing model size.