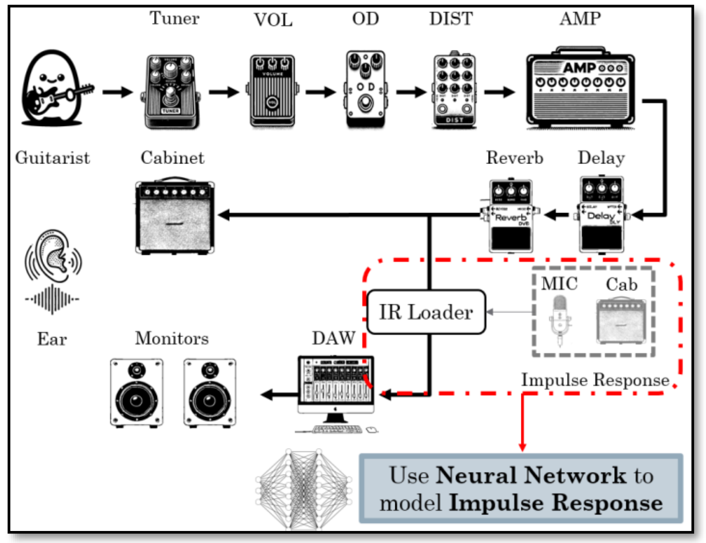
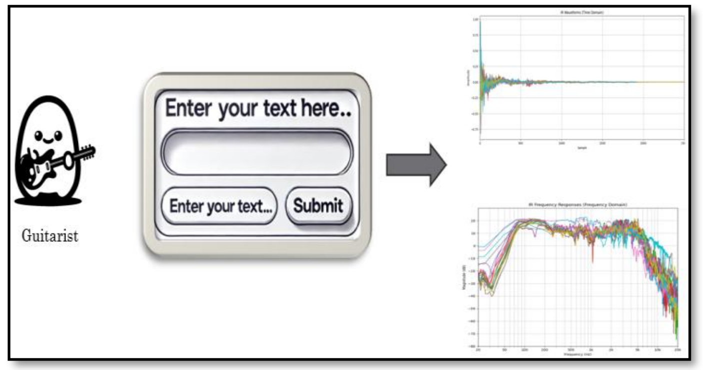
GTR-IR: GENERATIVE TEXT-CONDITIONED MODEL FOR GUITAR CABINET IMPULSERESPONSE

Jia-Chang Yang

Introduction

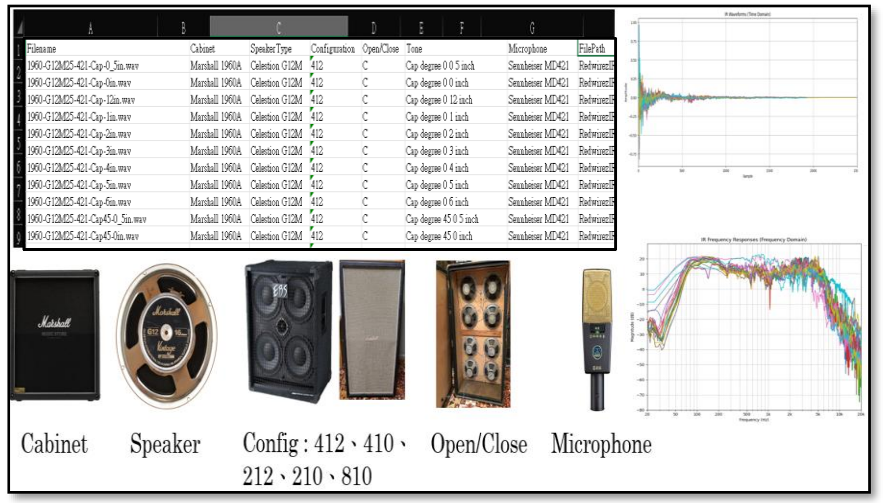
- Guitar tone depends on **effects, amplifiers, cabinets type, microphone type** and especially **speaker type**.
- Finding optimal cabinet IR is difficult.
- **Impulse Response (IR) convolution** enables cabinet simulation.
- Deep learning (WaveNet, RNN, etc.) has succeeded in amp modeling, but **cabinet IR generation remains underexplored**.
- **This study:** Propose a **Text-to-IR** model to generates IR directly from **text descriptions**

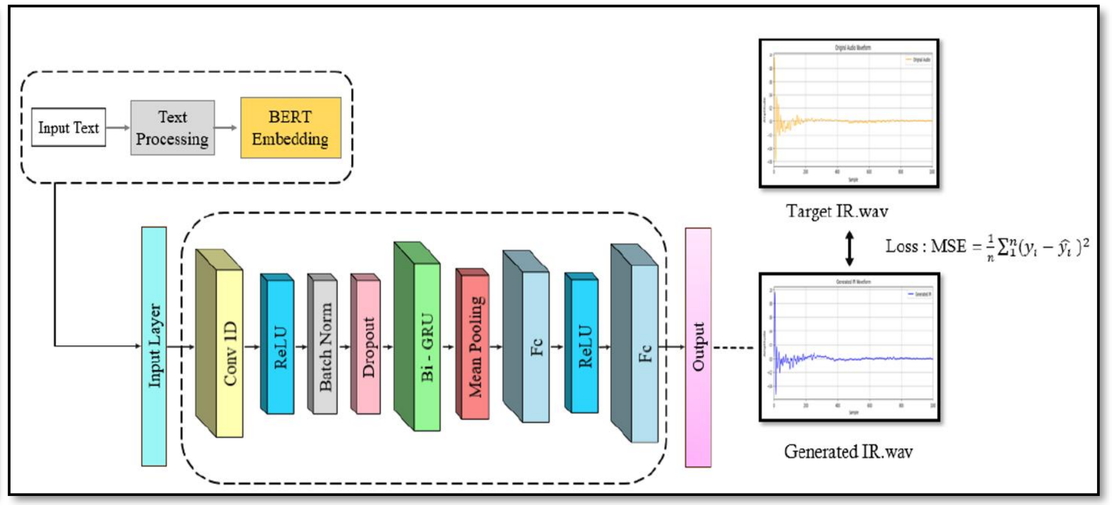




Method

- Input: Text descriptions (cabinet, speaker, configuration, microphone, tone).
- Text Encoder: BERT → semantic embeddings (768-dim).
- Model Architecture: CNN-BiGRU network.Conv1D (64 channel, kernel=3) → BiGRU (128x2) → pooling → FC → IR (2400 samples ≈ 50 ms, Sample Rate at 48 kHz).
- Training: MSE loss, AdamW optimizer.
- Dataset: 46,564 IRs, include. Redwirez Free IR Pack + custom recordings
- Novelty: Direct Text → IR mapping, semantics-driven cabinet simulation.

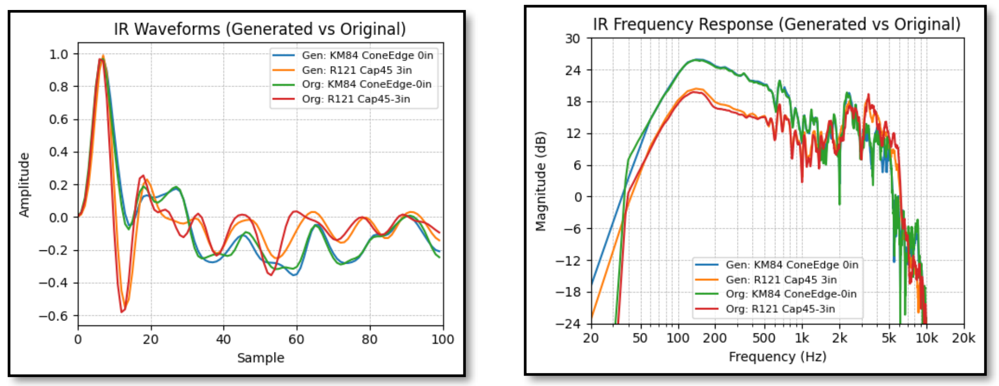




Results

Example generated IR results :

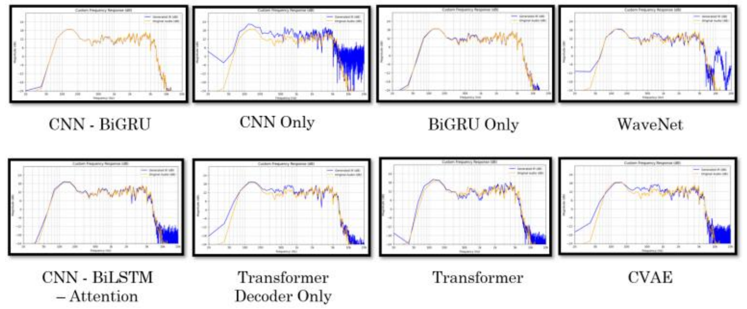
- Comparison of generated and original IR using Neumann KM84 placed on-axis at the cone edge (0 inch distance) and Royer R121 positioned 3 inch from the cap with a 45° off-axis angle. (a) Time-domain waveform. (b) Frequency response.



(a) (b)

Model Architecture Comparison :

- Compared with 8 alternative architectures, our CNN-BiGRU achieves the lowest MSE (1.78x10⁻⁷), highest spectral correlation (0.99998), and lowest residual error (0.66%).correlation (**0.99998**), and lowest residual error (**0.66%**).
- Confirms that combining CNN + BiGRU outperforms individual or alternative architectures.



Model	MSE	Spectral Corr.	Residual Energy Ratio (%)
CNN-BiGRU	**1.78e-07**	**0.99998**	**6.64e-03**
CNN Only	2.92e-05	0.99690	1.15
BiGRU Only	7.83e-07	0.99991	3.17e-02
LSTM	1.50e-06	0.99986	5.51e-02
CNN-BiLSTM-Attention	4.75e-07	0.99995	1.85e-02
Transformer Decoder Only	1.16e-04	0.98741	4.60
Transformer	2.07e-06	0.99982	7.99e-02
WaveNet	1.38e-04	0.97563	5.32
CVAE	1.67e-04	0.98541	6.56

Baseline Comparison :

- Against the MLP-based cabinet IR Baseline model, our approach achieves: MSE ↓ 96.5%, ESR ↓ 39%, PSDE ↓ 99.996% ,MSC ↑ 8.1%, Max correlation held steady .
- Results consistent across sample rates (44.1k – 96k Hz).

Fs	Model	MSE	ESR (%)	Max Xcorr	PSDE	MSC
44.1k	Baseline	4.240e-04	5.87	0.984	3.250e-03	0.880000
	Ours	1.487e-05	3.25	0.984	1.029e-07	0.999997
48k	Baseline	2.050e-04	2.88	0.988	3.000e-03	0.930000
	Ours	2.213e-05	5.24	0.973	8.607e-08	0.999985
88.2k	Baseline	2.550e-04	3.57	0.997	1.650e-03	0.961000
	Ours	1.944e-06	0.86	0.996	7.517e-08	0.999996
96k	Baseline	2.710e-04	3.79	0.998	1.500e-03	0.927000
	Ours	9.079e-07	0.44	0.998	6.943e-08	0.999997