

CCI-MOBILE: ENVIRONMENT-SPECIFIC SPEECH ENHANCEMENT WITH COCHLEAR IMPLANT LISTENERS USING CONVOLUTIONAL NEURAL NETWORK



Nursadul Mamun, Soheil Khorram, John H. L. Hansen
 Center for Robust Speech Systems (CRSS): Cochlear Implant Processing (CI) Lab
 Erik Jonsson School of Engineering & Computer Science,
 University of Texas at Dallas, Richardson, Texas, U.S.A.
 (Nursadul.Mamun, Soheil.Khorram, John.Hansen)@utdallas.edu



1. INTRODUCTION

- ◆ **Observation:** Cochlear Implant (CI) recipients enjoy near-to-normal speech intelligibility (SI) in quiet conditions [1].
- ◆ **Challenge:** CI recipients generally experience reduced speech intelligibility in environmentally rich spaces [2].
- ◆ **CNN-based Speech Enhancement (SE):** considers local temporal-spectral structures of speech signals and effectively dis-integrates the clean signal and noise from noisy input.
- ◆ **CNN network** ensures a powerful model capable of estimating the nonlinear mapping between noisy and clean speech.
- ◆ **Proposed networks:** Consider 3 Solutions: (i) vanilla CNN, (ii) spectral-subtraction style CNN (SS-CNN), (iii) Wiener-style CNN (Wiener-CNN).

2. METHODS

- ◆ **Method #1:** Speech Enhancement in a cochlear filter-bank space.
- ◆ **Method #2:** Leveraging the CNN to extract both stationary and non-stationary components of the environmental acoustics and speech

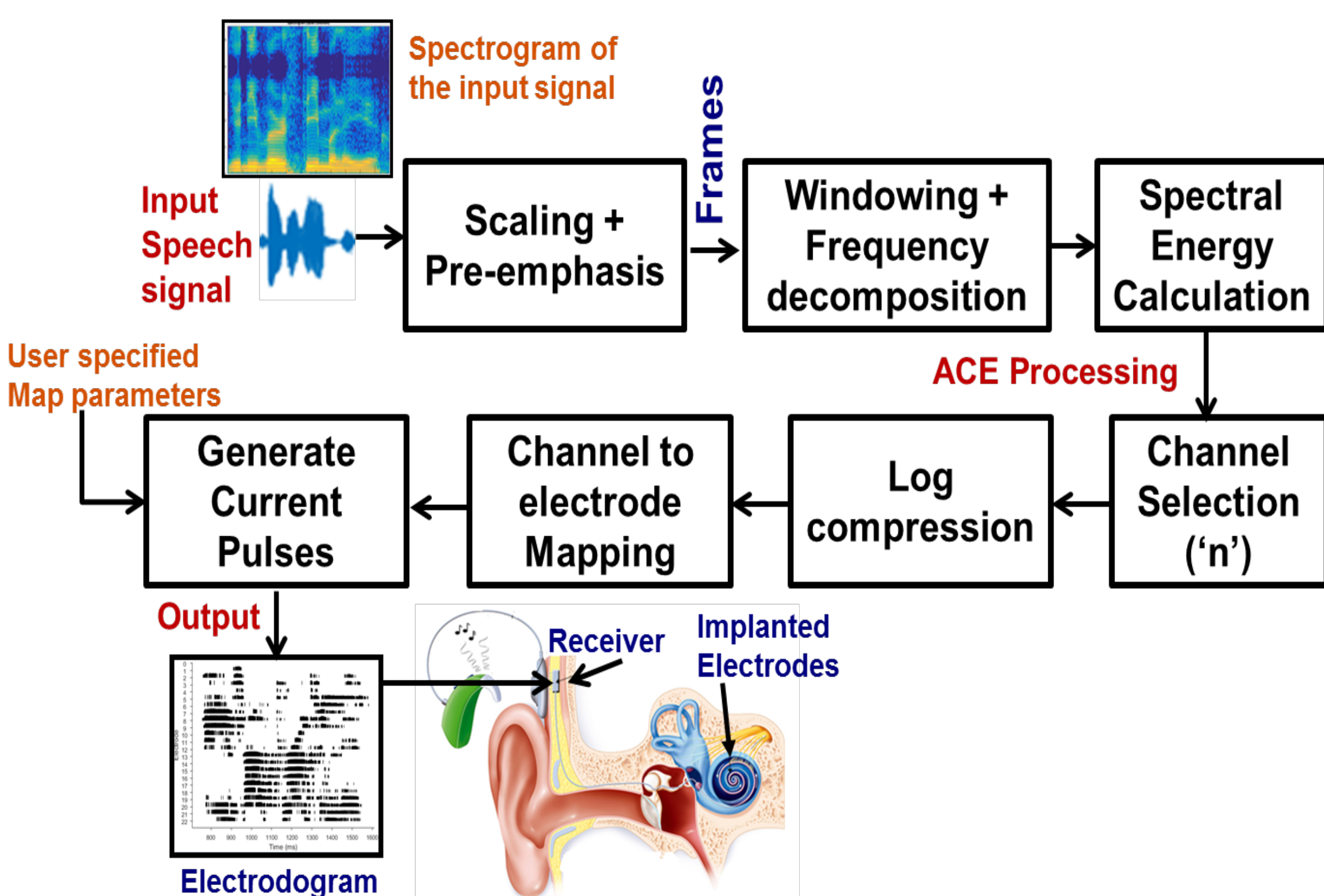
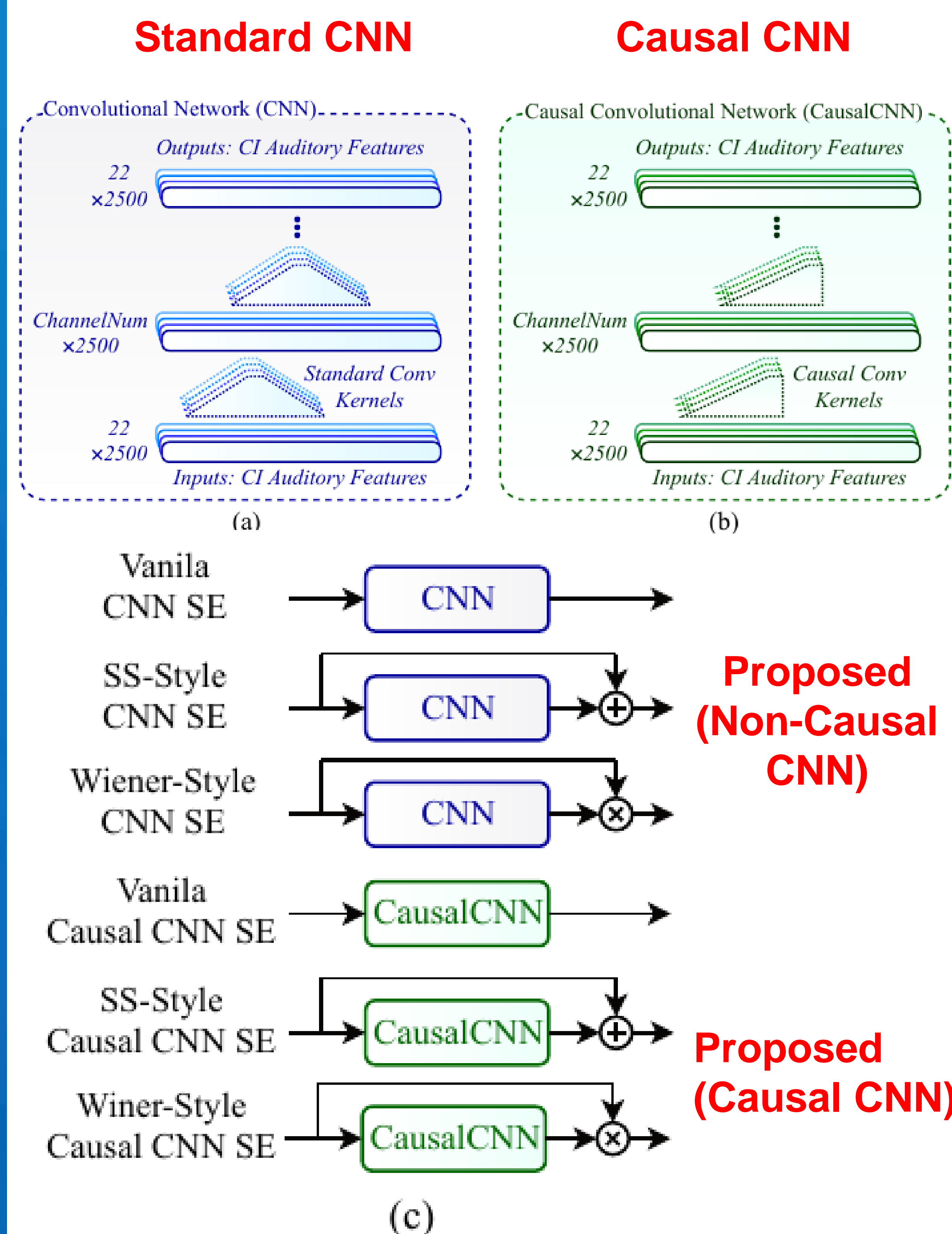


Fig.1. Basic block diagram of cochlear Implant signal processing pipeline.

3. METHODS cont.



CNN Parameters:

- ◆ **Training/Testing Sets:** 3150/1575 utterances
- ◆ **Convolutional Layer:** 7
- ◆ **Epochs:** 300. **Kernels:** 65
- ◆ **Activation Function:** 'tanh', 'linear'

4. RESULTS

Test Corpora:

- ◆ **Databases:** UT-Drive (TIMIT sentences).
- ◆ **SNR:** -10, 0, and 5 dB.
- ◆ **Noise:** Speech shape, Car.
- ◆ **Car Types:** Mitsubishi Galant (2002) and Nissan-Sentra (2008)

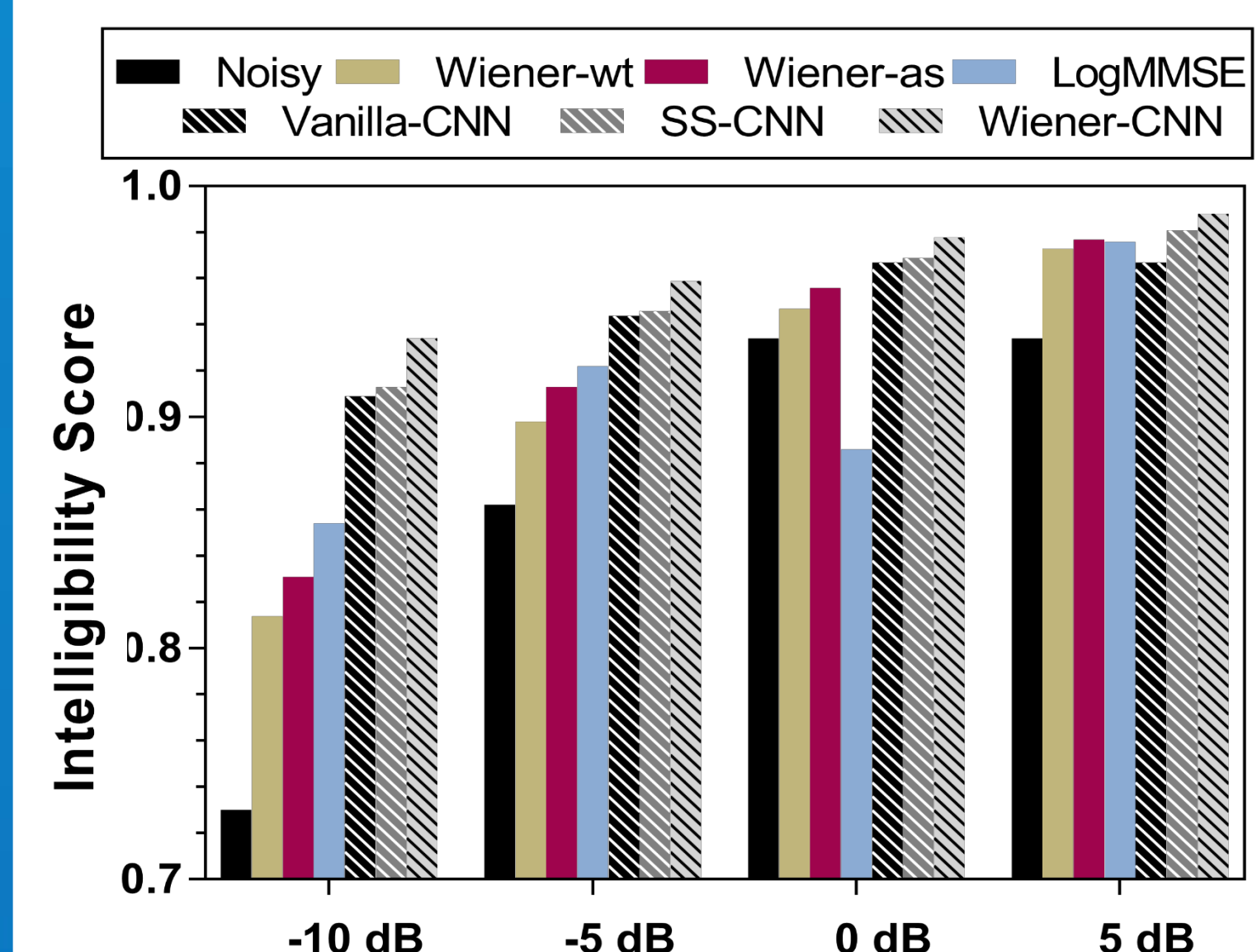
ECM metric is used to calculate the objective speech intelligibility score

Subjective Study:

- ◆ **Implant users:** 1 unilateral and 1 bilateral.
- ◆ **Databases:** 120 pairs (sentences) (pair of noisy and enhanced signal)
- ◆ **SNR:** 5 and 10 dB
- ◆ **Noise:** Speech shape

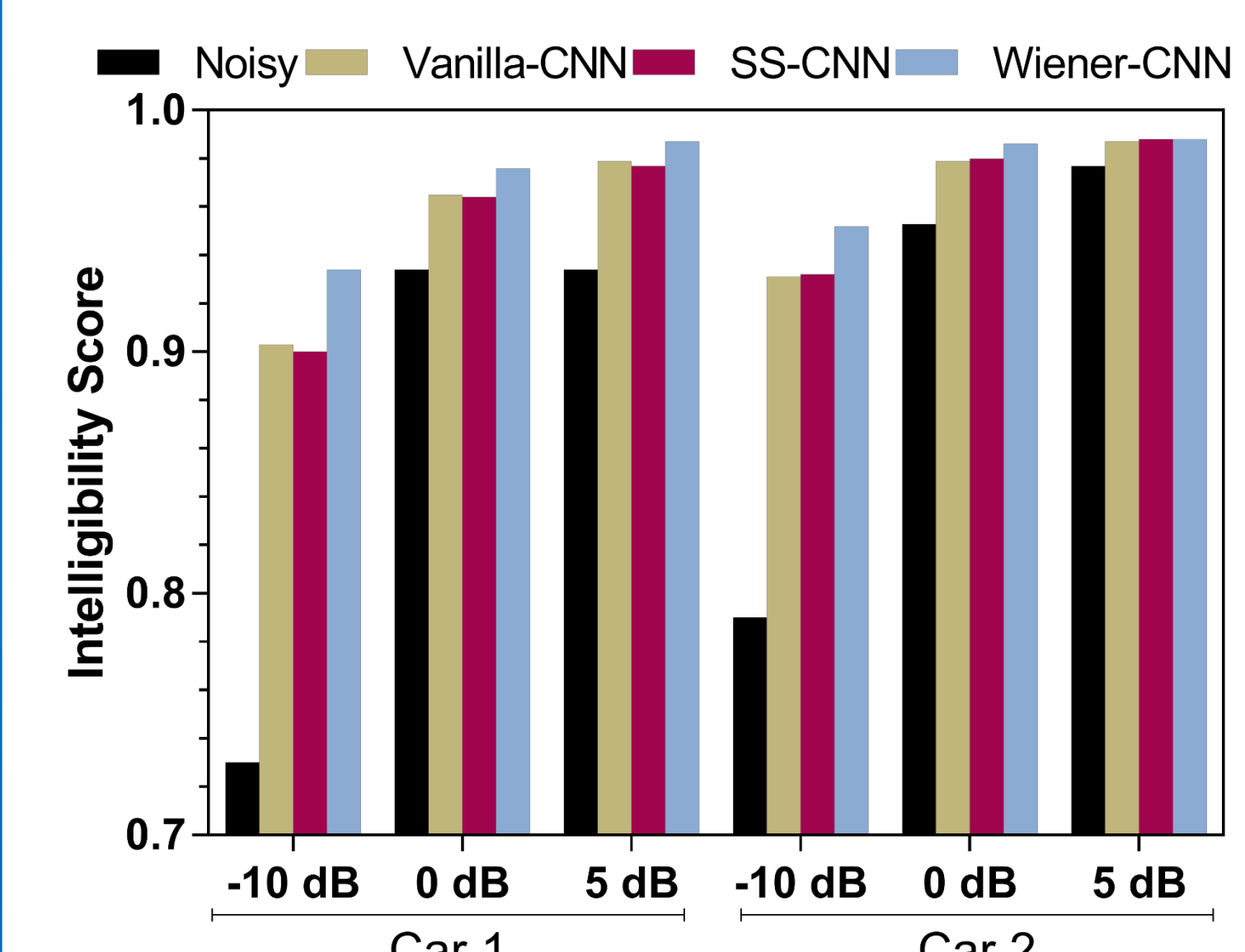
4. RESULTS cont.

Intelligibility scores (Non-causal SE):



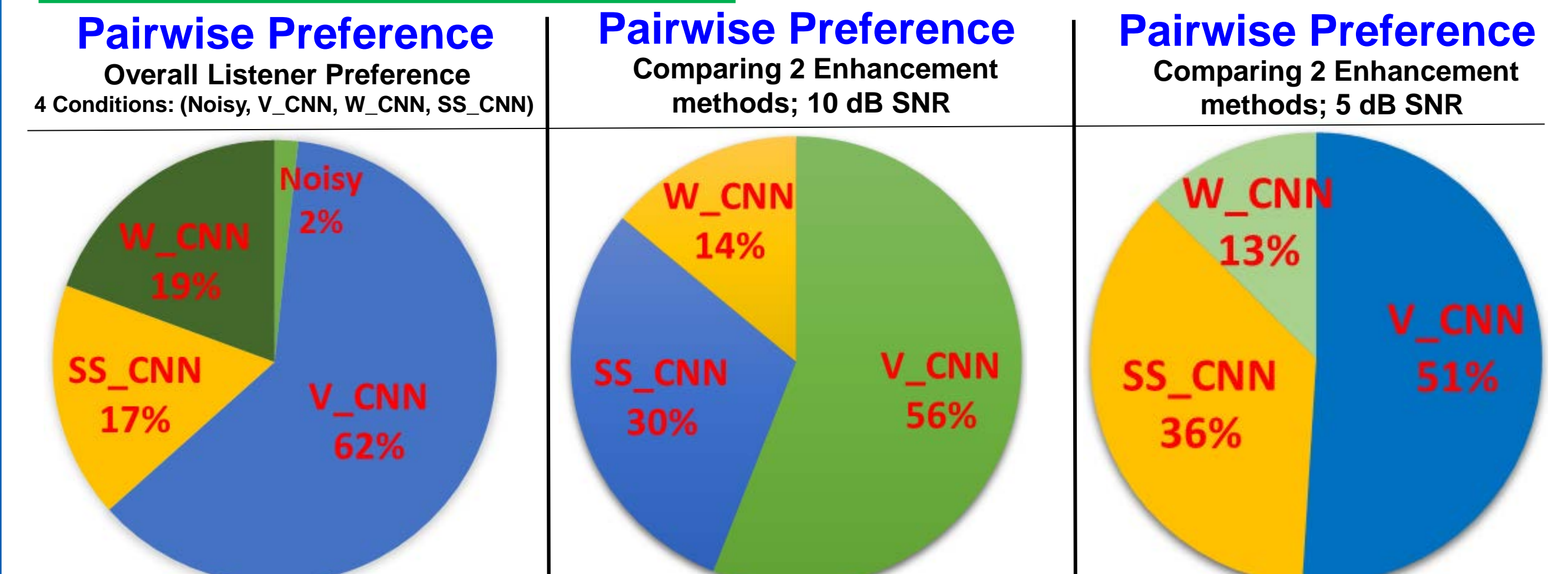
- ◆ 3 proposed system and 3 baseline systems
- ◆ Wiener-CNN outperform other systems under car noise
- ◆ Outperform baseline algorithms.

Intelligibility scores (Causal SE):



- ◆ 3 proposed system systems
- ◆ Significantly improve SE in CI features domain
- ◆ Causal Wiener-CNN (SE) algorithms outperform others

Paired Preference Test:



Vanilla-CNN based SE outperforms all other methods for speech shaped condition.

5. CONCLUSIONS

- ◆ Proposed algorithms shown to successfully enhanced signals under noisy conditions for CI users.
- ◆ **Causal Wiener-CNN** achieves the highest performance vs. all 3 methods.
- ◆ CI subjects preferred enhanced sentences (98%) versus noisy sentences (2%).
- ◆ CI subjects preferred **Vanilla-CNN** based enhanced signal vs. others for speech shaped noise (5 & 10dB SNR).

REFERENCES

- 1) F.G. Zeng; S. Rebscher (2008). *IEEE review in biomedical engineering* 1, 115.
- 2) L. M. Friesen, R. V. Shannon, (2001) *Journal ARO* 110(2):1150.
- 3) B. C. Moore (2008) *Journal of the Association for Research in Otolaryngology* 9(4):399.
- 4) J.R. Dubno; A.R. Horwitz; J.B. Ahlstrom. (2005) *Journal of the ASA* 118, 923.
- 5) G.A. Studebaker; R.L. Sherbecoe, D.M. McDaniel, C.A. Gwaltney (1999) *JASA* 106, 2111.