1505: CCI-MOBILE: COMPARATIVE ANALYSIS OF CNN-BASED MODELS VS HUMAN SOUND RECOGNITION AMONG CI AND NH SUBJECTS

Ram Charan M Chandra Shekar, John H.L Hansen

Center for Robust Speech Systems (CRSS), Cochlear Implant Processing Lab
Erik Jonsson School of Engineering & Computer Science,
University of Texas at Dallas, Richardson, Texas, U.S.A.
(ramcharan.chandrashekar, john.hansen)@utdallas.edu

Cochlear Implant Laboratory

1. INTRODUCTION

Environmental Sound (ES) Perception

- Important for hearing-specific quality of life (QoL) among CI users
- Limitations of ES perception: (i) limited number of studies, (ii) large reported variability in outcomes due to experimental factors, (iii) lack of CI subject recruitment (limited access to subjects)
- However, many studies report no substantial improvement in ES perception for post implantation (CI users) when compared to CI candidates

Objective

- Evaluate ES perception for CI users and assess the performance against a CNN-based ES identification model **[1]**

2. METHODS

Sound Battery/Audio Dataset

- ESC 50 Database: 50 environmental sound classes, 5 categories, 2000 sound samples, 40 samples per class

Cochlear Implant Signal Processing

- CCi-MOBILE: Uses an ‘*n*’-of-’*m*’ strategy (ACE, Cochlear Corp.), generates electric stimuli using a standard CI user MAP (200/100 MCL/THR for all 22 channels, ‘*n*’ = 8)
- Braecker Vocoder: Uses a two-sided, exponentially decaying function with 2/3rd Gammatone filter bank, generates electric stimuli to synthesized (or auralized) stimuli

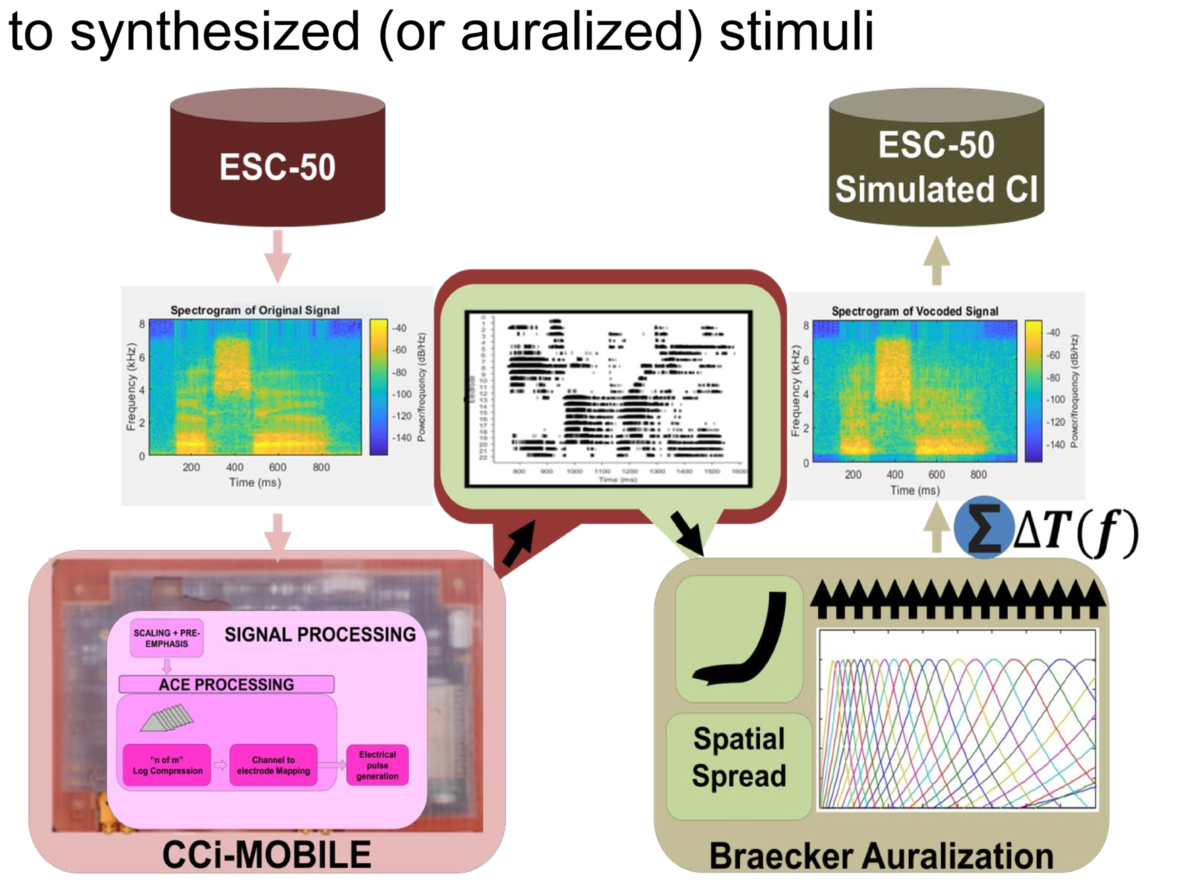


Fig. 1. Signal processing block diagram of simulated CI conditions using the CCi-MOBILE and Braecker Auralization process.

REFERENCES

[1] Shekar, R.C., Belitz, C., & Hansen, J.H.L. (2021). IEEE SLT-2021: Spoken Language Technology Workshop, pp. 728-733.

3. COMPARATIVE ANALYSIS

CNN Features for Sound Representation

- Fully convolutional VGG-based CNN
- Input: 128-band mel-spectrograms
- Output: 1024 dim sound representations
- 8 Layers: 2/1 conv layers | batch norm | max pooling | ReLU/Sigmoid
- Classifier: SVM – Grid Search Algorithm

NH Machine Model Framework

- CNN frameworks for CNN+SVM, CI stimuli, NH stimuli, respectively

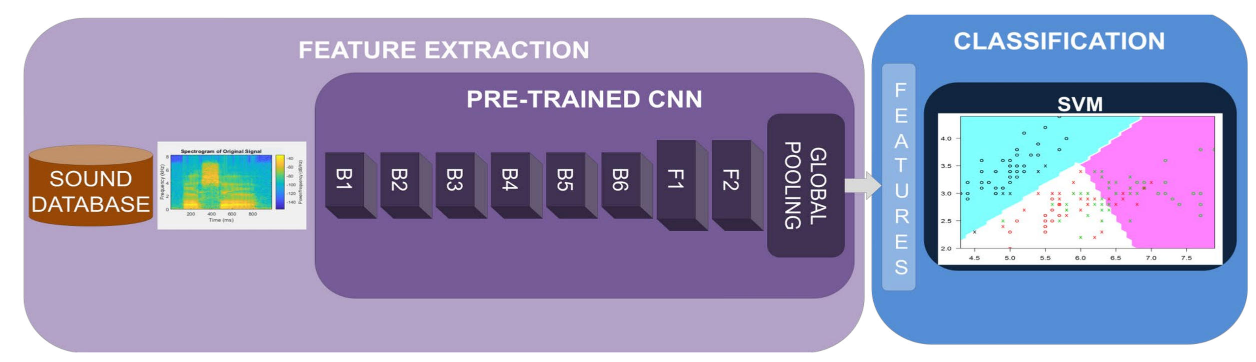


Fig. 2. Framework for sound classification using pre-trained CNN and SVM classifier.

Fig. 3. Sound classification framework for CI listeners.

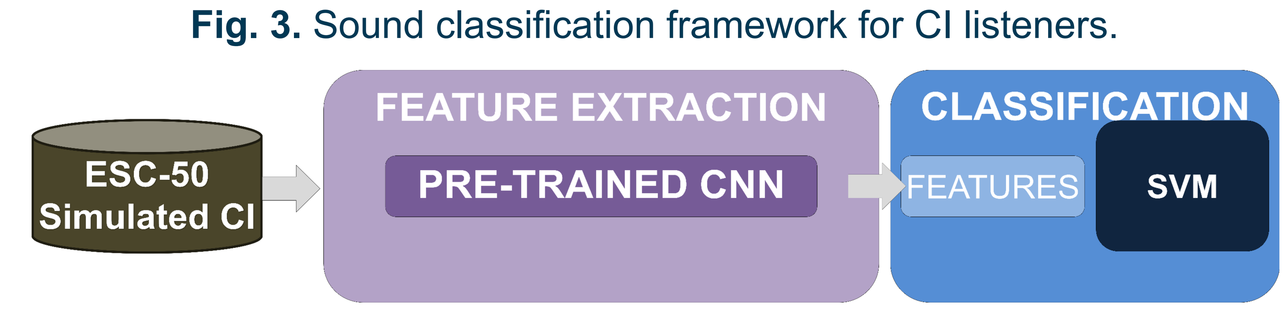


Fig. 4. Sound classification framework in simulated CI conditions for NH listeners.

5. CONCLUSIONS

- 2 CI users, 4,000 NH judgements **[1]**
- NH and CI models recorded high mean classification accuracies compare to the listener evaluation
- Higher variation in CI listeners (SD=29.43%) compared to NH listeners (SD=14.25%)
- Identification scores of CI users were found to be correlated (33.02%) with CNN model; more than three times as compared to NH listeners (10.07%)
- CI users resulted in higher identification scores for animal sound category (>50%) with the exception of insects sound class
- Higher variation in sound identification among CI listeners could be attributed to other factors such as (but not limited to) familiarity, cognition, and memory

4. RESULTS

Mean Classification Accuracy

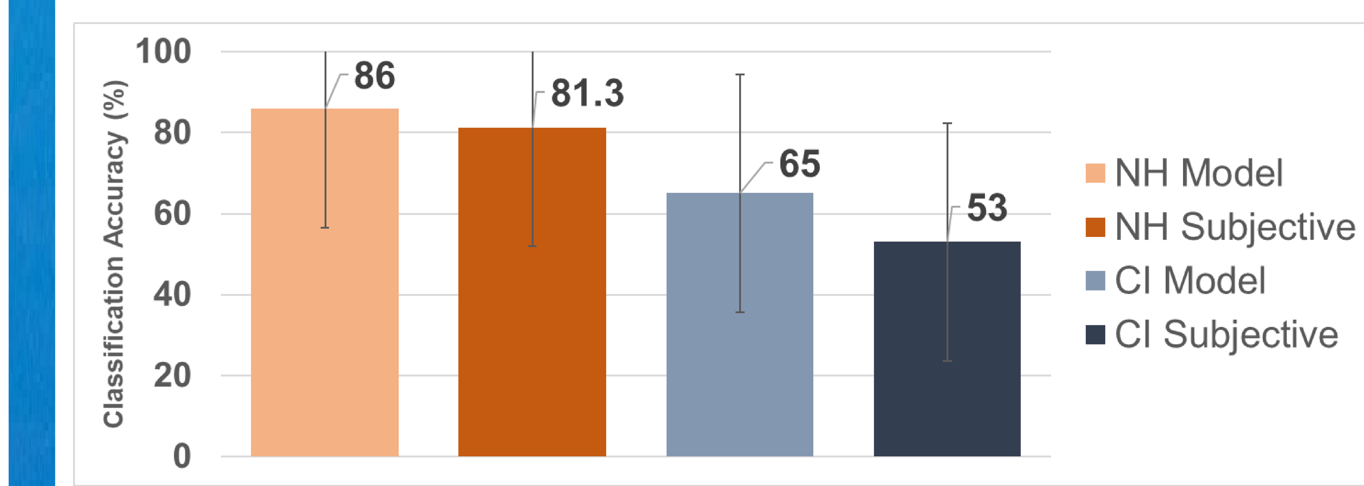
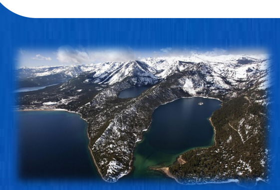


Fig. 5. Mean sound classification accuracy for model/subjective groups.

Identification Accuracy

Sound Classes	Normal Hearing (NH)		Cochlear Implant (CI)	
	Model	Subjective	Model	Subjective
Animal Sound Category				
DOG	88%	100%	75%	75%
ROOSTER	100%	71%	100%	100%
PIG	88%	89%	75%	50%
COW	75%	94%	75%	50%
FROG	88%	75%	88%	75%
CAT	50%	88%	50%	50%
HEN	88%	77%	62%	50%
INSECTS	100%	99%	50%	0%
SHEEP	100%	95%	88%	75%
CROW	75%	77%	62%	50%
Nature Sound Category				
RAIN	62%	78%	62%	25%
SEA WAVES	75%	68%	62%	25%
CRACKLING FIRE	100%	63%	100%	75%
CRICKETS	88%	52%	75%	50%
BIRD CHIRPING	100%	84%	62%	100%
WATER DROPS	100%	92%	25%	75%
WIND	100%	46%	88%	25%
POURING WATER	100%	75%	75%	100%
TOLIET FLUSH	88%	88%	62%	50%
THUNDERSTORM	100%	85%	62%	100%
Human Non-Speech Sound Category				
CRYING BABY	100%	99%	88%	100%
SNEEZING	100%	88%	62%	75%
CLAPPING	100%	92%	100%	25%
BREATHING	100%	89%	88%	50%
COUGHING	62%	94%	62%	75%
FOOTSTEPS	75%	83%	50%	75%
LAUGHING	62%	97%	25%	25%
BRUSHING TEETH	88%	89%	88%	50%
SNORING	100%	84%	100%	75%
DRINKING WATER	88%	80%	38%	25%
Interior Sound Category				
DOOR KNOCK	100%	90%	100%	100%
MOUSE CLICK	62%	65%	50%	25%
KEYBOARD CLICKS	100%	83%	100%	25%
DOOR CREAK	50%	90%	12%	50%
CAN OPENING	88%	80%	75%	50%
WASHING MACHINE	75%	34%	25%	25%
VACUUM CLEANER	100%	58%	88%	25%
ALARM CLARM	88%	92%	62%	50%
CLOCK TICK	88%	89%	62%	75%
GLASS BREAKING	100%	99%	62%	75%
Exterior Sound Category				
HELICOPTER	62%	64%	25%	0%
CHAINSAW	88%	83%	38%	75%
SIREN	88%	93%	62%	50%
CAR HORN	75%	90%	25%	25%
ENGINE	75%	82%	50%	25%
TRAIN	88%	67%	62%	0%
CHURCH BELLS	100%	95%	75%	75%
AIRPLANE	38%	68%	38%	25%
FIREWORKS	100%	68%	75%	0%
HAND SAW	88%	90%	75%	100%

Fig. 6. Identification accuracies per category for NH/CI model/listeners.



2021 Conference on Implantable Auditory Prostheses
Lake Tahoe, CA (Virtual), July 12 – 16, 2021

This work was supported by the grant R01DC016839 from the National Institutes of Health, National Institute on Deafness, and Other Communication Disorders