Deciphering Everyday Hidden Hearing Loss Using Statistical and Machine Learning Methods

Zhou Long,

Grade 11, Westmount Charter School

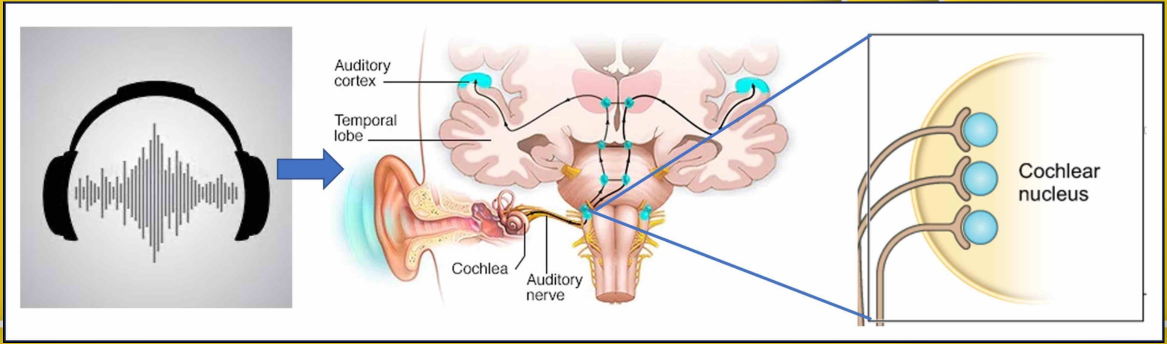
Supervised by Dr. Jun Yan (MD, Professor), University of Calgary

Background

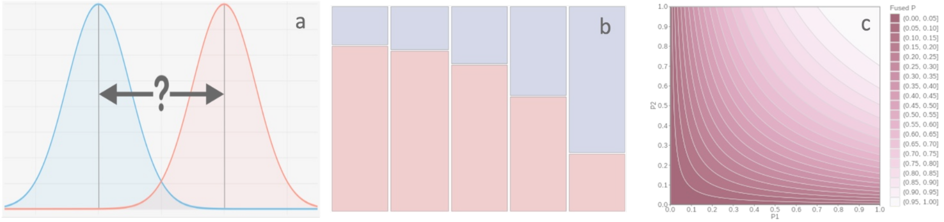
- More than 1.5 Billion people, or 20% of the global population suffer from hearing loss (WHO). This number is expected to increase to over 2.5 Billion by 2050.
- Between 12% and 15% of these cases are caused by hidden hearing loss. That is over 200 million people.

- Hidden hearing loss not assessable by conventional hearing tests.
- Additionally, it is especially dangerous due to possible damage from “safe” sounds.
- Damage occurs in the form of atypical neuronal activities in the auditory system of the brain.

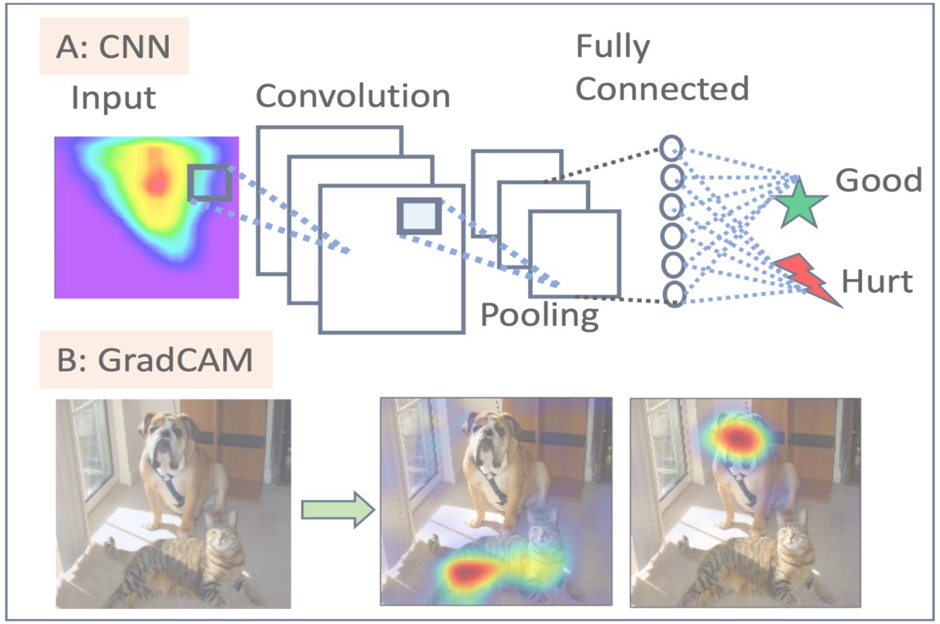
Question: How can we quantify and detect hidden hearing loss?



Pictures from the internet



$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{2}{n}}} \qquad T \equiv \sum_{i=1}^{k} t_i(N_{1i}R_2 - N_{2i}R_1) \qquad X^2_{2k} \sim -2\sum_{i=1}^{k}\log(p_i)$$



Aims of the Project

1. Using rigorous statistical tests (upper panel) to **reveal** that hidden hearing loss exists in the form of atypical neuronal response.
2. Using machine learning techniques (lower panel) to **detect** hidden hearing loss and decipher the **specific spectrum** of amplitude and frequency for a clinical diagnostics.

Contributions:

- The experimental data was collected by Ms. Wenyue Xue (a PhD student in Dr. Jun Yan's group).
- My contributions are the statistical and computational works achieving the two Aims. Code and Software for the analysis are in my GitHub: https://github.com/ZhouLongCoding/sound_waves

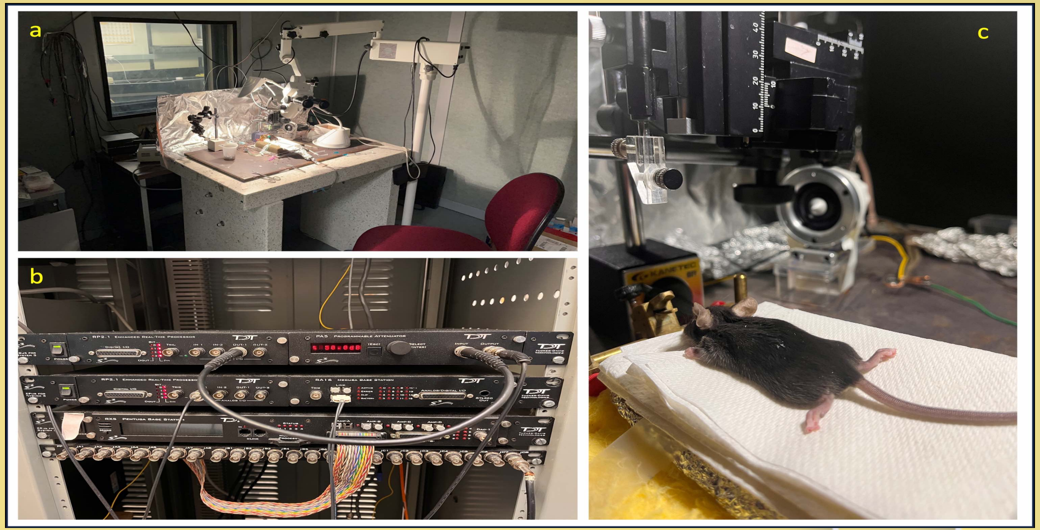
Methods (1.1): Experiments & Data processing

Mouse experiments

a. A mouse is fixed on the bench, exposing to a pre-specified pure-tone sound.

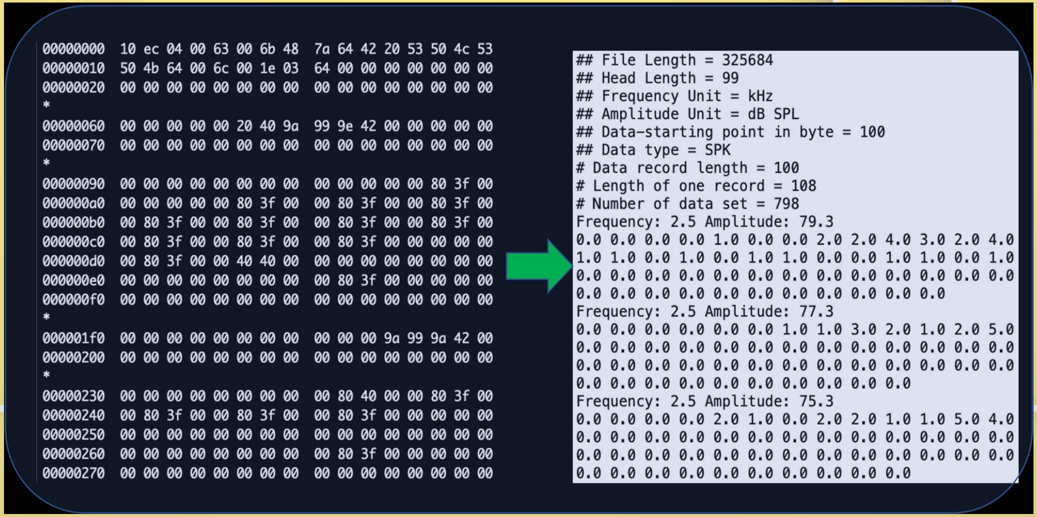
b. Pre- and post- the responses of auditory midbrain neurons to sound stimuli are recorded by a specific equipment and assessed via conventional methods.

c. An enlarged view of the animal experiment scenario



Data Processing

1. Conversion from binary to texts
2. Smoothing the data



In total, the input data configuration:

- 18 mice are experimented.
- 36 Amplitudes 21 frequencies, at 100 time points are assessed for each mice at 2 conditions (pre- and post-)
- Total number of the data points = 2,721,600

Mouse pictures from Ms. Wenyue Xue, Dr. Jun Yan's lab at the University of Calgary.

Methods (1.2): Feature Extraction for Statistical Tests

The **Raw data** after processing are depicted 4D data (Neuronal Firing Rate with respect to three dimensions: **Frequency**, **Amplitude**, **Time**) in the **Upper panel.** Different Time points are shown in **a – h**. In each sub-panel Frequency is shown in x-axis, and Amplitude is shown in y-axis, the colour indicates strength of the neuronal firing rate.

Feature extractions are conducted (**Lower pane**):

Frequency domain (fixing both Amplitude and Time)

- **Best frequency**, the Frequency with the highest firing rate (**Lower panel a**);
- **Bandwidth**, the frequency range that has firing-rate substantially differ from zero (**Lower panel b**)

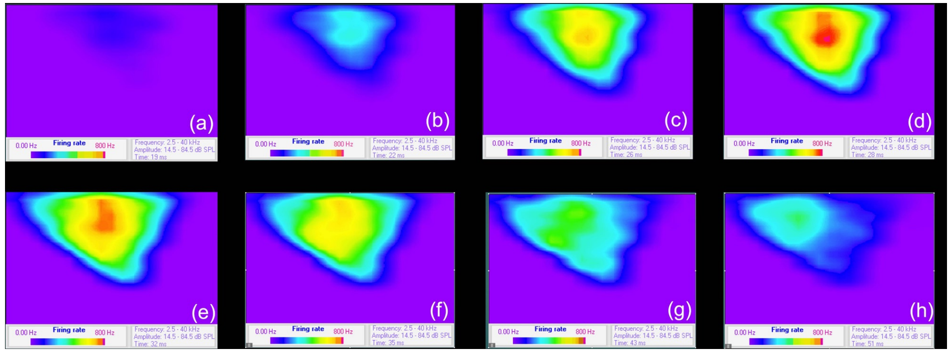
Amplitude domain (fixing both Frequency and Time)

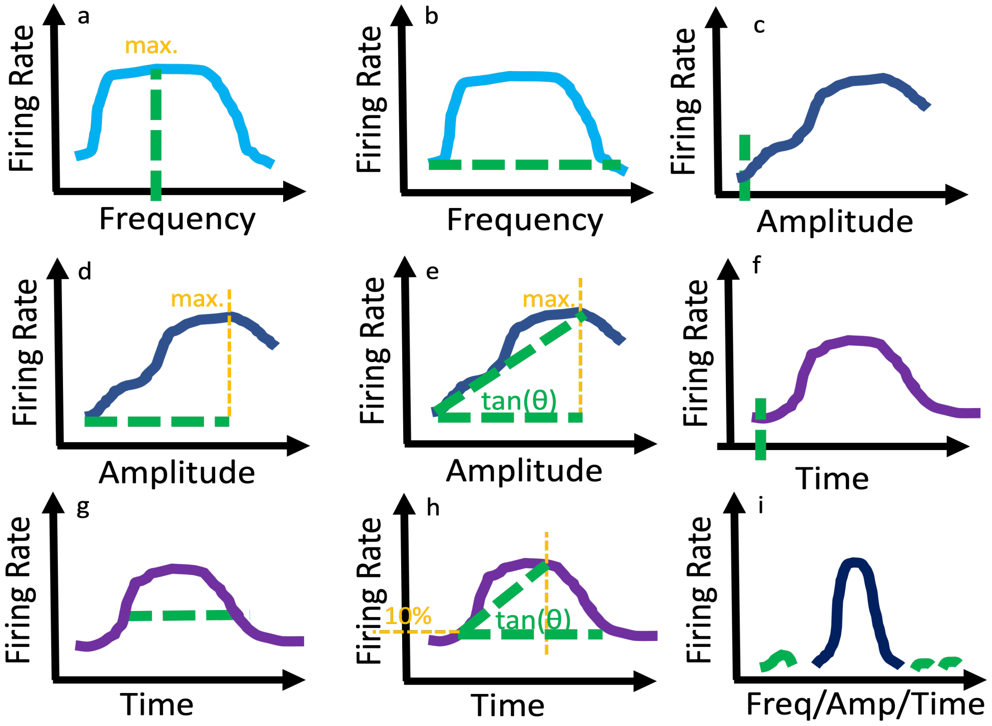
- **Threshold**, the lowest response amplitude (**Lower panel c**);
- **Dynamic range**, the amplitude difference between lowest firing-rate and the turn point (after which the increase of firing-rate slows down), which is approximately maximal the second derivative of the firing rate curve (**Lower panel d**);
- **Slope of dynamic range**, the distance in firing rate divided by the dynamic range described above (**Lower panel e**).

Time domain (fixing both Frequency and Amplitude)

- **Latency**, the time point when a response starts (**Lower panel f**);
- **50% Duration**, the period when firing rate is over 50% of the maximum firing-rate of each response (**Lower panel g**);
- **Rising slope**, the slope from the 10% of the maximum firing-rate to the maximum firing-rate (**Lower panel h**).

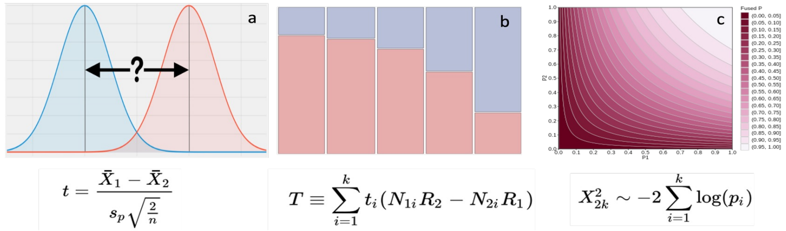
Noise/Signal Ratio to ((**Lower panel i**) for all the three domains.

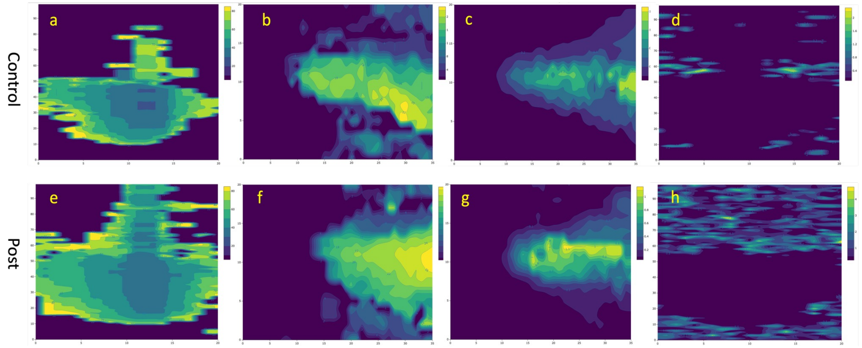


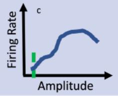
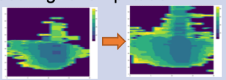
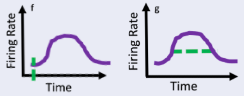
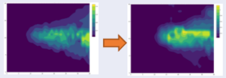
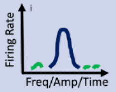
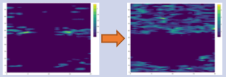


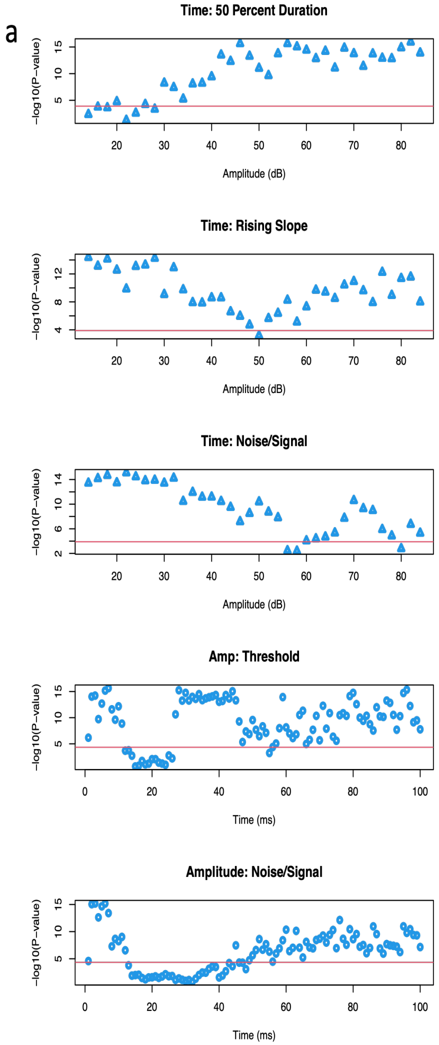
Results 1: Hidden hearing loss exists! (Statistical Tests + Interpretations)

- T-test and Trend-test (**Upper-left panel, a** and **b**) are used to assess the significant level of effects. Fisher method (**Upper-left panel, c**) is used to combine multiple mice experiments (Pictures illustrating tests from internet)
- Significance level in T-tests (**Right panel**) for five features.
- The above features are illustrated as contour graphs (**Mid-left panel, a-d:** pre-exposure; **e-h:** poser exposure; **a,e:** Amplitude Threshold**; b,f:** Time 50 duration; **c.g:** Time rising slope; **d,h:** Amplitude Noise/Signal ratio)
- The illustrated features are annotated using physiological interpretations (**Lower-left panel, a – c**) (Interpretation provided by Dr. Jun Yan)





Features	Neuronal Response	Practical Problem
Threshold 	Need higher amplitude. 	**Unable to hear soft sounds**
50% Duration **Rising slope** 	Weaker response to low stimuli; stronger response to high stimuli. 	**Insensitive to low signals. However, overreaction to loud sounds**
Noise/Signal Ratio 	Fake signals are fired in contrast to genuine signals. 	**Can hear; but non-existent signals are also present**



Methods (Aim 2): Identifying specific spectrums for clinical tests

Goal:

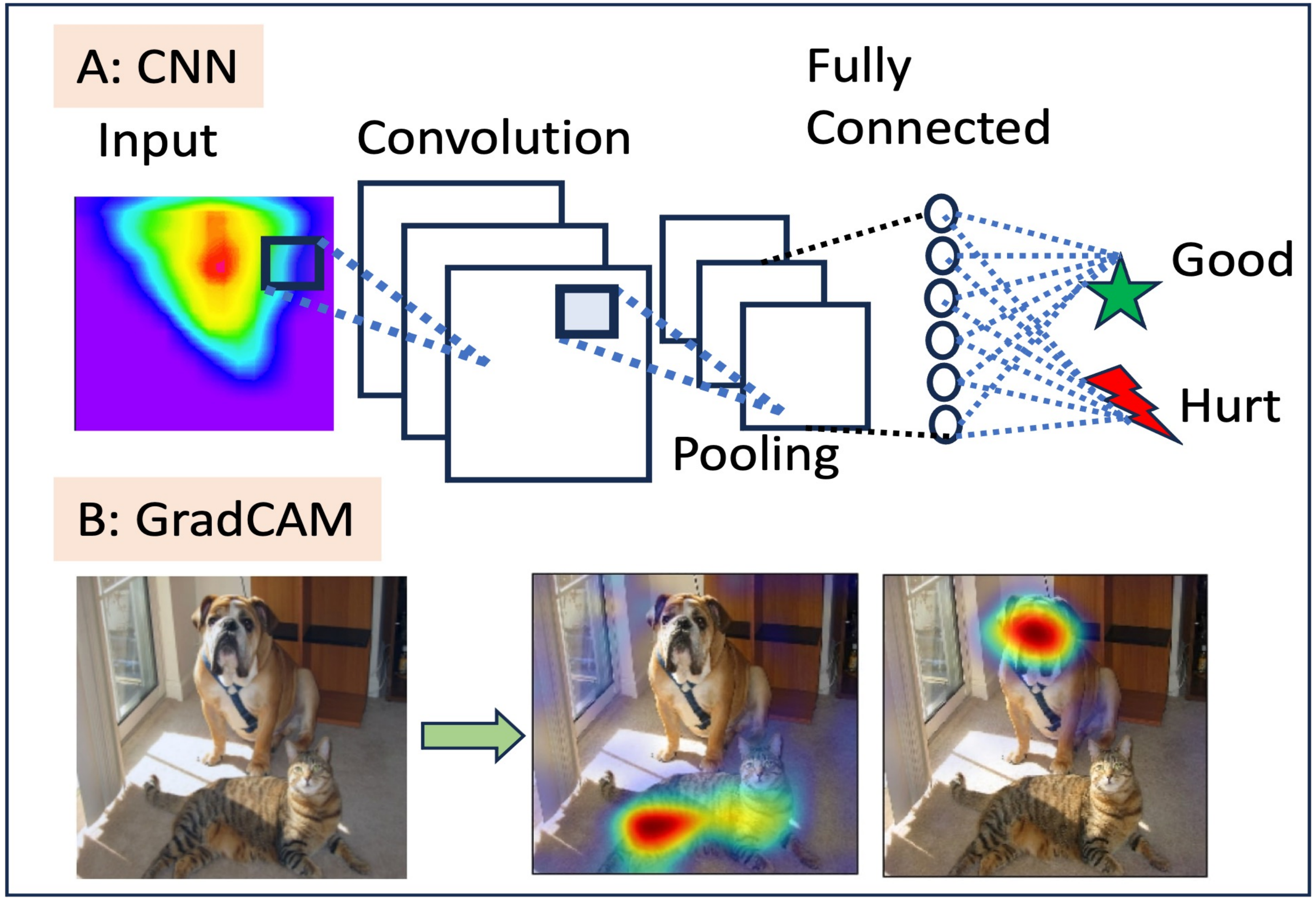
- Detect and identify specific amplitude and frequency that best distinguish pre-exposure and -post exposure.

The analytic pipeline:

- Convolutional neural network is used to train a classifier distinguishing pre- and post-exposure using input data as images. This allows detection of hidden hearing loss, which has been previously impossible.
- eXplainable AI (particularly GradCAM) is used to identify the spectrums (amplitude and frequency) that best distinguish pre- and post exposures.

Overfitting control:

- De-noising the input data
- Cross-Validation
- Early stopping



Cat/Dog image from the GradCAM paper: Selvaraju *et al.*, ICCV 2017

Hyper-parameters & Training for Convolutional Neural Network

Network configurations:

- Kernel size = (3, 3)
- Pool size = (2,2)
- Dense layer number of nodes = 128
- Number of layers =3

Training parameters:

- Epochs = 20
- Batch Size = 12
- Optimizer = ‘adam’
- Learning rate = 0.001
- Kernel initializer = HeNormal()

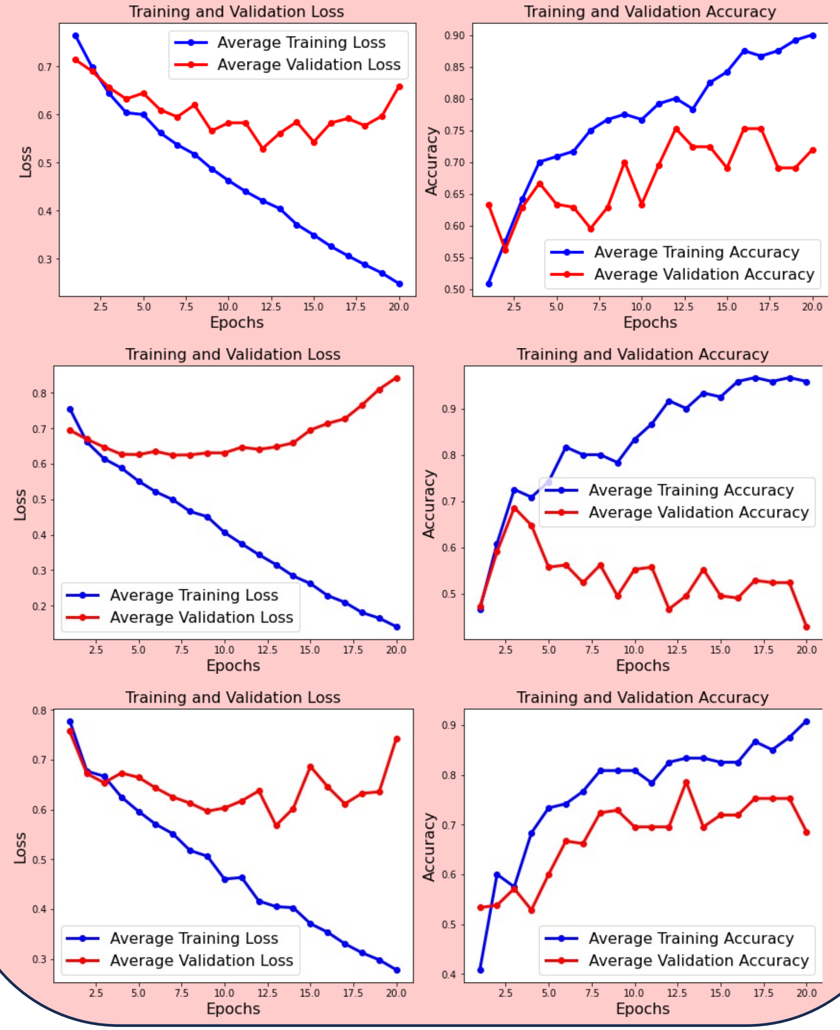
Data processing:

- De-noise: Median filter
- Mean normalization:
  - X’ = (X – mean(X))/(max(X)-min(X))

Model summary

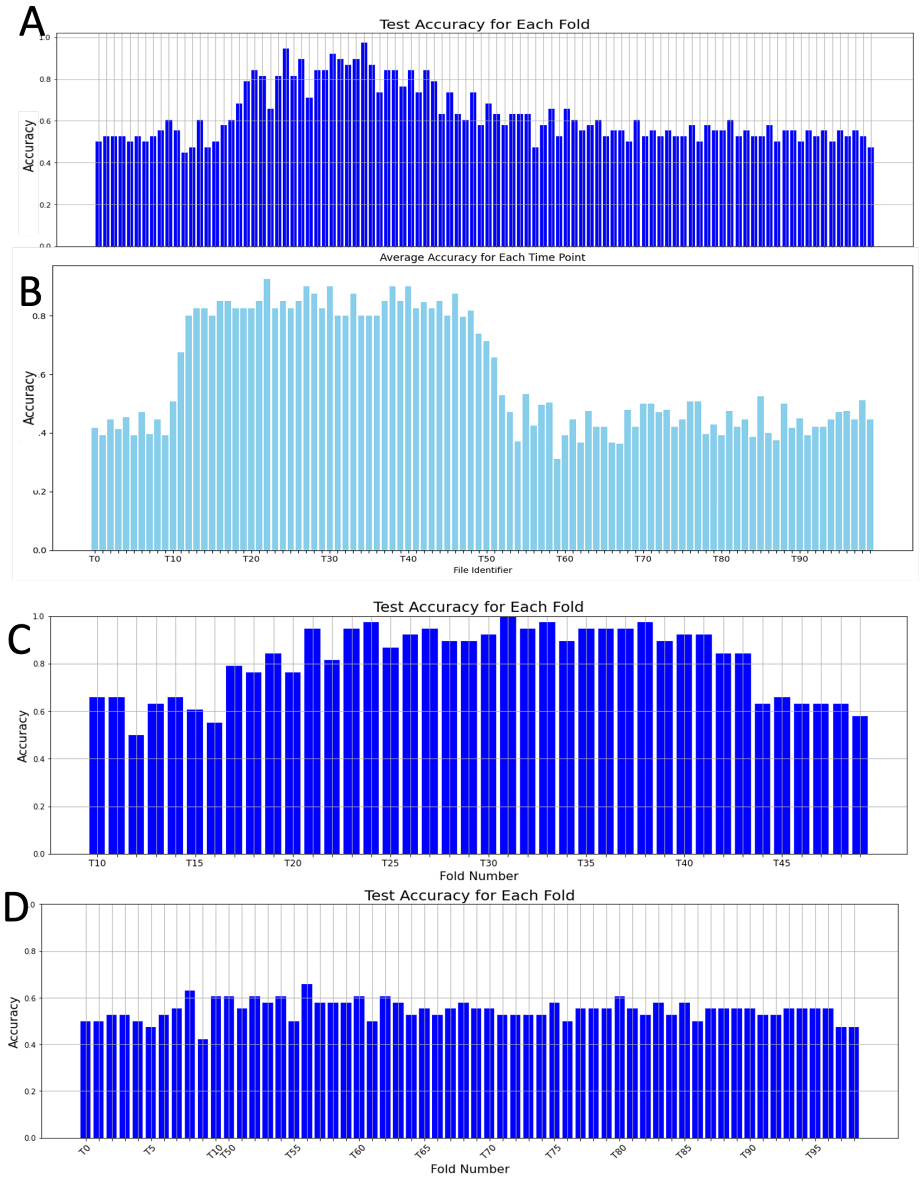
```
Layer (type)                  Output Shape            Param #
==============================================================
conv1 (Conv2D)                (None, 33, 19, 32)      320

MaxPooling1 (MaxPooling2D)    (None, 16, 9, 32)       0

conv2 (Conv2D)                (None, 14, 7, 64)       18496

MaxPooling2 (MaxPooling2D)    (None, 7, 3, 64)        0

conv3 (Conv2D)                (None, 5, 1, 128)       73856

flatten_4 (Flatten)           (None, 640)             0

Dense (Dense)                 (None, 128)             82048

Output (Dense)                (None, 2)               258

==============================================================
Total params: 174978 (683.51 KB)
Trainable params: 174978 (683.51 KB)
Non-trainable params: 0 (0.00 Byte)
```

Loss & Accuracy graphs:



Predictive Accuracy & Overfitting Control

- Data in Time point 0-9 and 50-99 are known (biologically) to be noisy.
- I used this prior knowledge to analyze potential overfitting in the training process and select the best model
- First, when trains all data jointly, the accuracy in the range of T=10 to T=49 indeed are better (**Panel A**).
- Second, when trained individually, the advantage of T=10 to T=49 are more pronounced (**Panel B**).
- Third, when analyzing the data jointly for T=10 to T=49 (**Panel C**) and T=0 to T=9 and T=50 to T=99 (**Panel D**), the performance of T=10 to T=49 are way better, replicating the patterns observed in **Panels A** and **B**.
- The above observation shows that the model is no overfitting (because of the poor data serve as controls) and we should use only T=10 to T=49 for further discoveries.



Results 2: eXplainable AI (GradCAM) reveals the final spectrum

GradCAM generates weights specifying importance of each spots on the image, which are aggregated to lead the final outcome:

For a potential clinic test, The best range and spectrum is **Amp = 46-70 dB**, **Freq = 8.7-13.2** KHz (with respect to the specific exposure)

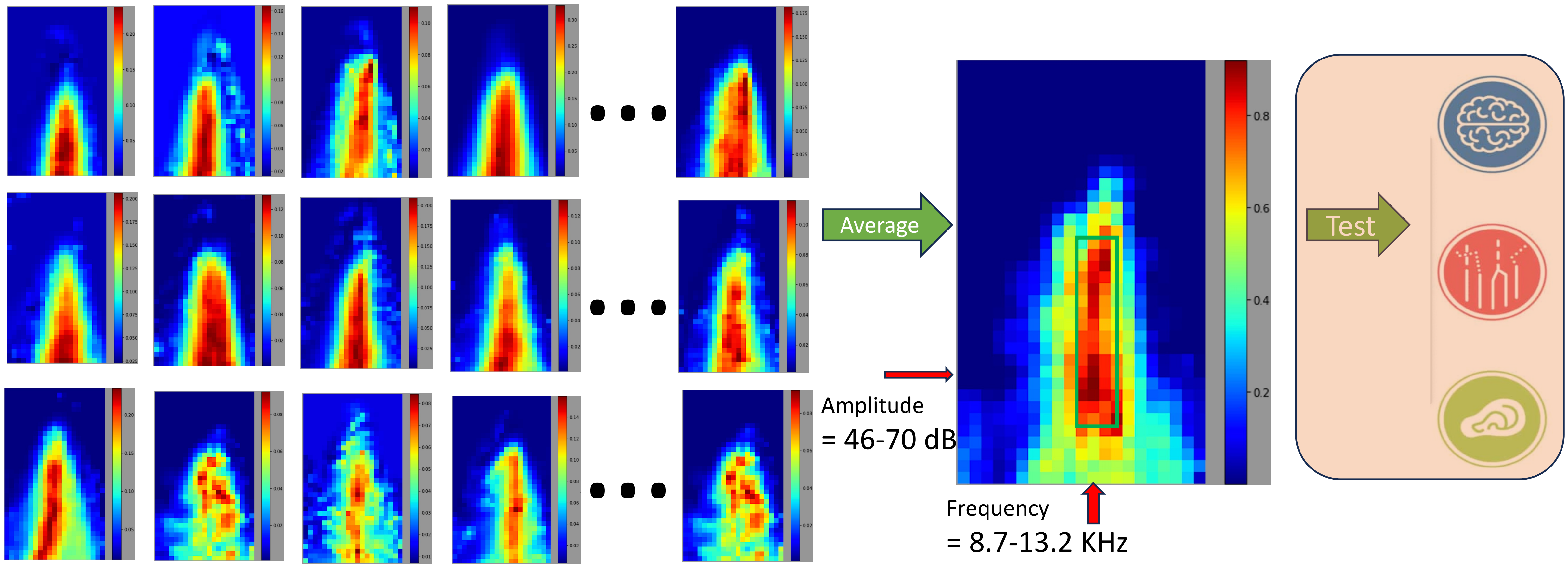
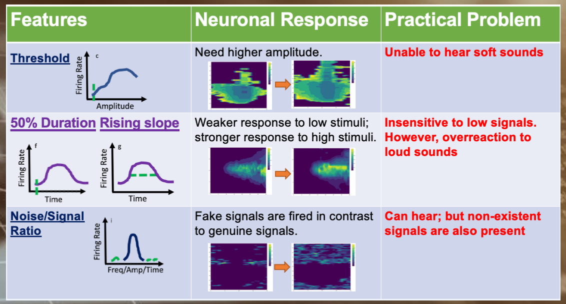
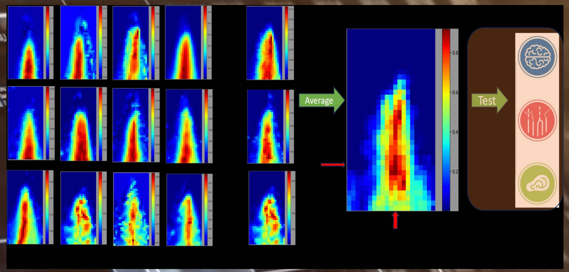
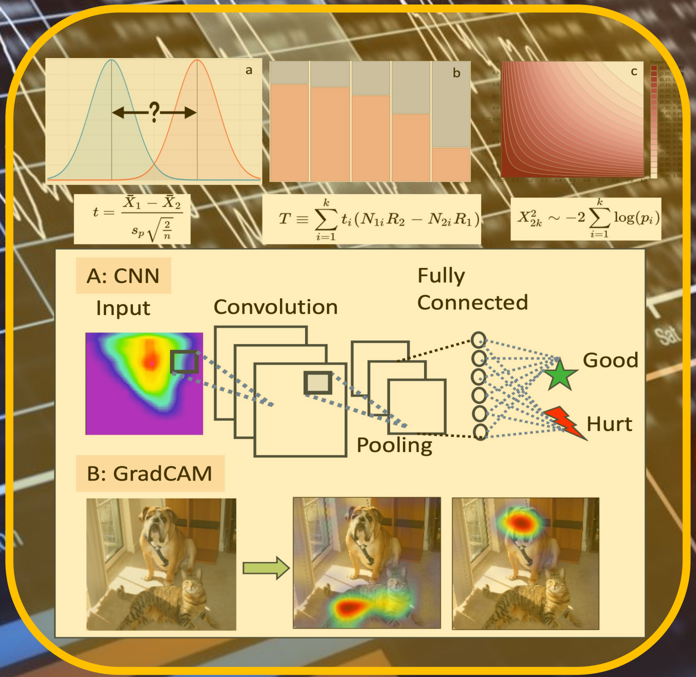


Image indicating clinical test from American Academy of Audiology



Features	Neuronal Response	Practical Problem
Threshold	Need higher amplitude.	Unable to hear soft sounds
50% Duration Rising slope	Weaker response to low stimuli; stronger response to high stimuli.	Insensitive to low signals. However, overreaction to loud sounds
Noise/Signal Ratio	Fake signals are fired in contrast to genuine signals.	Can hear; but non-existent signals are also present





Conclusion and Take Home Message

- Hidden hearing loss is significant and can be caused by safe sounds, shown by the statistical tests (Results 1)
- Trained a convolutional neural network to detect hidden hearing loss exists in a sample.
- Identified the specific range of Amplitude/Frequency that is the most informative in distinguishing pre- and post-exposure, laying the path for clinical tests. (Results 2)

WESTMOUNT CHARTER SCHOOL

UNIVERSITY OF CALGARY

References & Acknowledgement

[1] M. Charles Liberman, “Hidden hearing loss: Primary neural degeneration in the noise-damaged and aging cochlea,” *Acoustical Science and Technology*, vol. 41, no. 1. 2020. doi: 10.1250/ast.41.59.

[2] Z. Long, “Revealing Hidden Hearing Loss Caused by ‘Safe’ Sounds,” 2023. [Online].

Available: https://github.com/ZhouLongCoding/sound_waves/blob/main/FinalWrittenReport_size_reduced.pdf

[3] C. M. Bishop, *Machine Learning and Pattern Recoginiton*. 2006.

[4] A. Barredo Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, 2020, doi: 10.1016/j.inffus.2019.12.012.

[5] R. Bishop, F. Qureshi, and J. Yan, “Age-related changes in neuronal receptive fields of primary auditory cortex in frequency, amplitude, and temporal domains,” *Hear Res*, vol. 420, 2022, doi: 10.1016/j.heares.2022.108504.

[6] P. G. Hoel, “Introduction to Mathematical Statistics.,” *Economica*, vol. 22, no. 86, 1955, doi: 10.2307/2626867.

[7] P. Armitage, “Tests for Linear Trends in Proportions and Frequencies,” *Biometrics*, vol. 11, no. 3, 1955, doi: 10.2307/3001775.

[8] R. A. Fisher, “Statistical Methods, Experimental Design, and Scientific Inference.,” *Biometrics*, vol. 47, no. 3, 1991, doi: 10.2307/2532685.

[9] R. Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization ” International Conference on Computer Vision (ICCV'17)

[10] A. Chaudhary, K. S. Chouhan, J. Gajrani, and B. Sharma, “Deep learning with PyTorch,” in *Machine Learning and Deep Learning in Real-Time Applications*, 2020. doi: 10.4018/978-1-7998-3095-5.ch003.

Dr. Jun Yan

Ms. Wenyue Xue

Campbell McLaurin Foundation

NSERC CRSNG