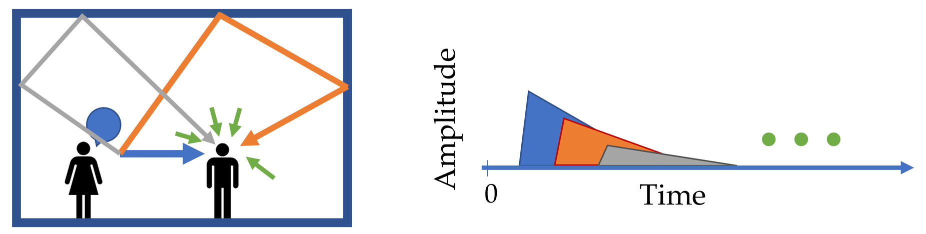
Reconstruction of Dry and Reverberant Speech from Ensemble Responses in the Auditory Midbrain

Oded Barzelay[1,2], Stephen David[3], Bertrand Delgutte[1,2]

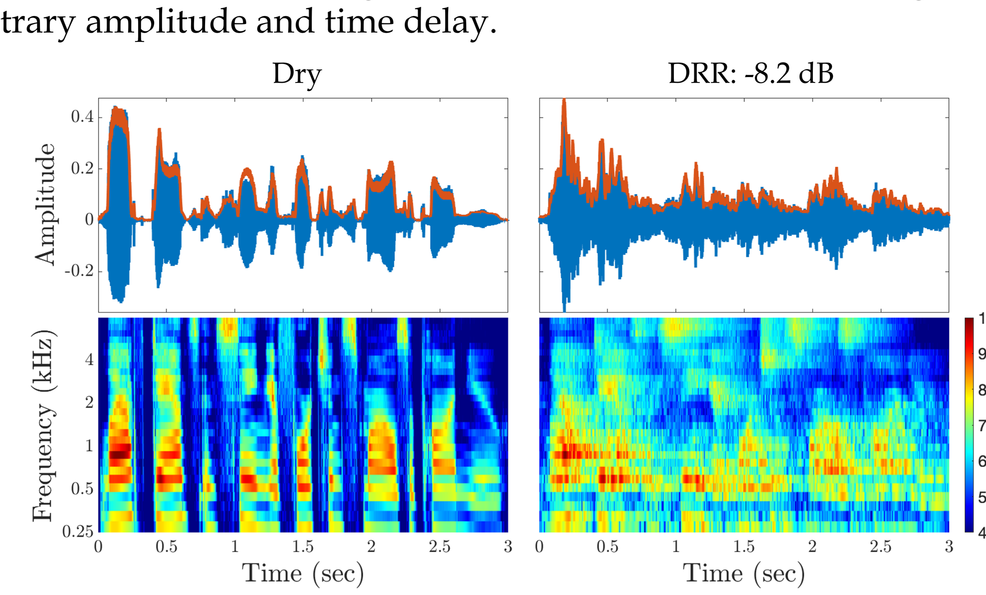
[1] Eaton-Peabody Laboratories, Massachusetts Eye & Ear Infirmary, [2] Dept. of Otolaryngology, Harvard Medical School, [3] Oregon Hearing Research Center, Oregon Health & Science University

Introduction

Reverberation is ubiquitous in everyday acoustic environments. It degrades both binaural cues and the envelope modulations of sounds and thus can severely degrade speech perception, especially in hearing-impaired individuals. Still, both humans and animals can accurately perceive reverberant stimuli in most everyday settings. Despite the importance of reverberation, the neural mechanisms that allow accurate localization and identification of sounds in reverberation are still unknown. Some studies from the auditory midbrain[1] and auditory cortex[2] have suggested the existence of neural mechanisms that partially compensate for the effects of reverberation. However, these studies were limited by their use of either highly simplified stimuli or rudimentary reverberation simulations.



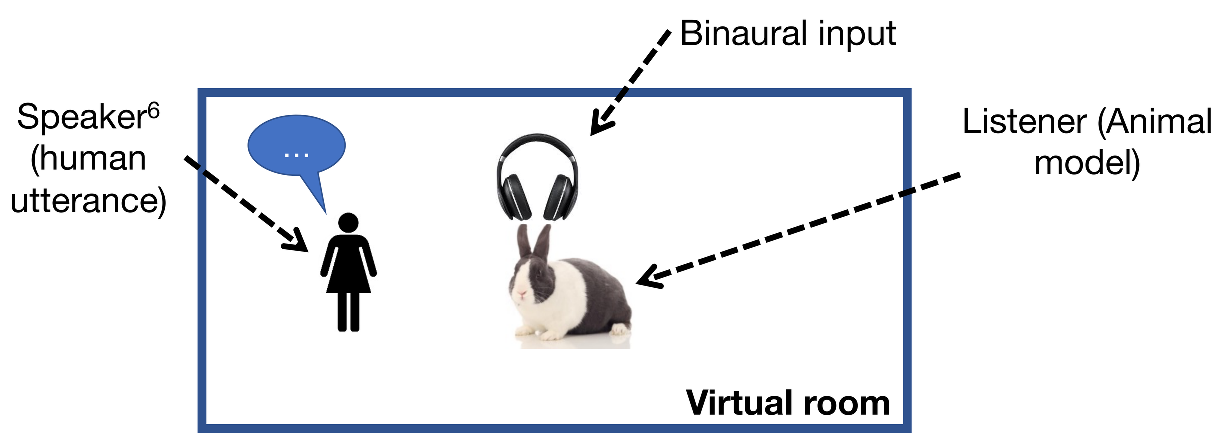
- In an echoic environment (room, corridor, etc.) reverberations accumulate with the direct stimulus signal. Each of these reverberated signals can have arbitrary amplitude and time delay.



- The overall effect of reverberation is thus the distortion of the stimulus envelopes.
- Neurons in the inferior colliculus (IC) are tuned to these envelopes.
- **Is there a mechanism that compensate for reverberation in the IC neurons?**

Methods

- **Animal model**: we record from the inferior colliculus (IC) of Unanesthetized Dutch-belted rabbit.



- **Binaural room impulse responses (BRIR)** are simulated using the room-image method[5] (Room size: $11m\times13m\times3m$).
- **Distance between sensors** is ~10 cm to match the maximum ITD of 300 μs for a typical rabbit[4].
- **There is no acoustic shadow** ("no head") between sensors ("ears") when simulating for the BRIR, so **there is only the effect of ITD without ILD**.
- The "listener" stands near the center of the room, and the speaker stands 1.5 m or 3.0 m in front of the listener.

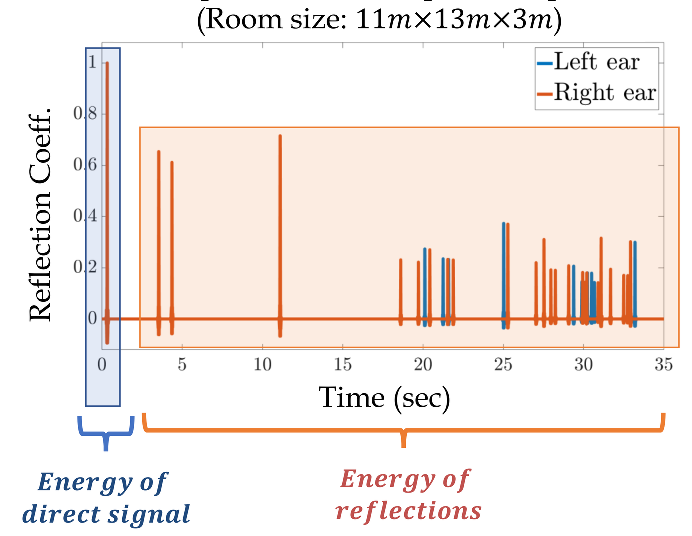
Direct-to-reverberant ratio (DRR)

Speech utterances[6] is convolved with room impulse response (RIR) to add reverberations. We used speech with no reverberation ("dry") and various degrees of simulated reverberations (direct-to-reverberant energy ratios ranging from 9.4 dB to -8.2 dB).

- Reverberation is measured with **Direct-to-reverberant ratio (DRR):**

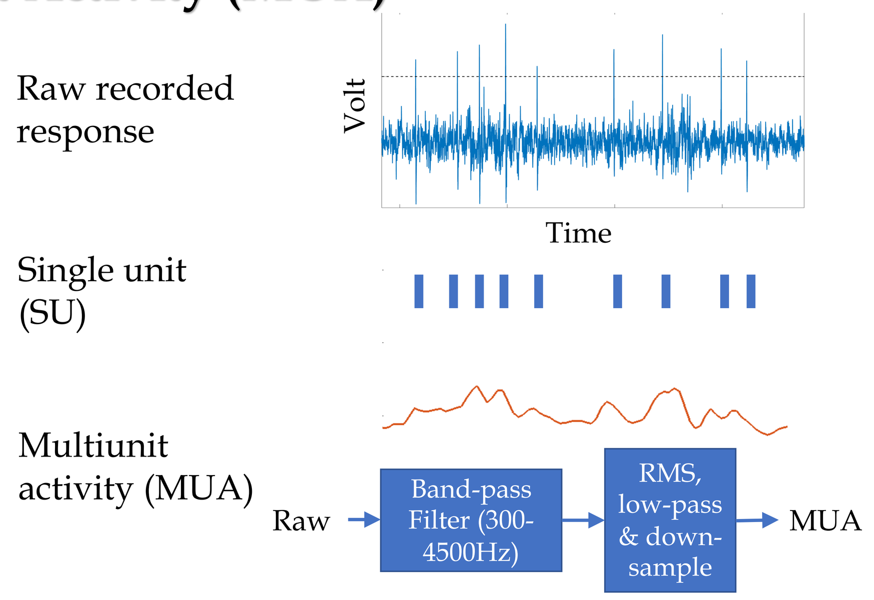
$$DRR = \frac{\text{Energy of direct signal}}{\text{Energy of reflections}}\ (\boldsymbol{dB})$$

Example of a Room Impulse Response (Room size: $11m\times13m\times3m$)



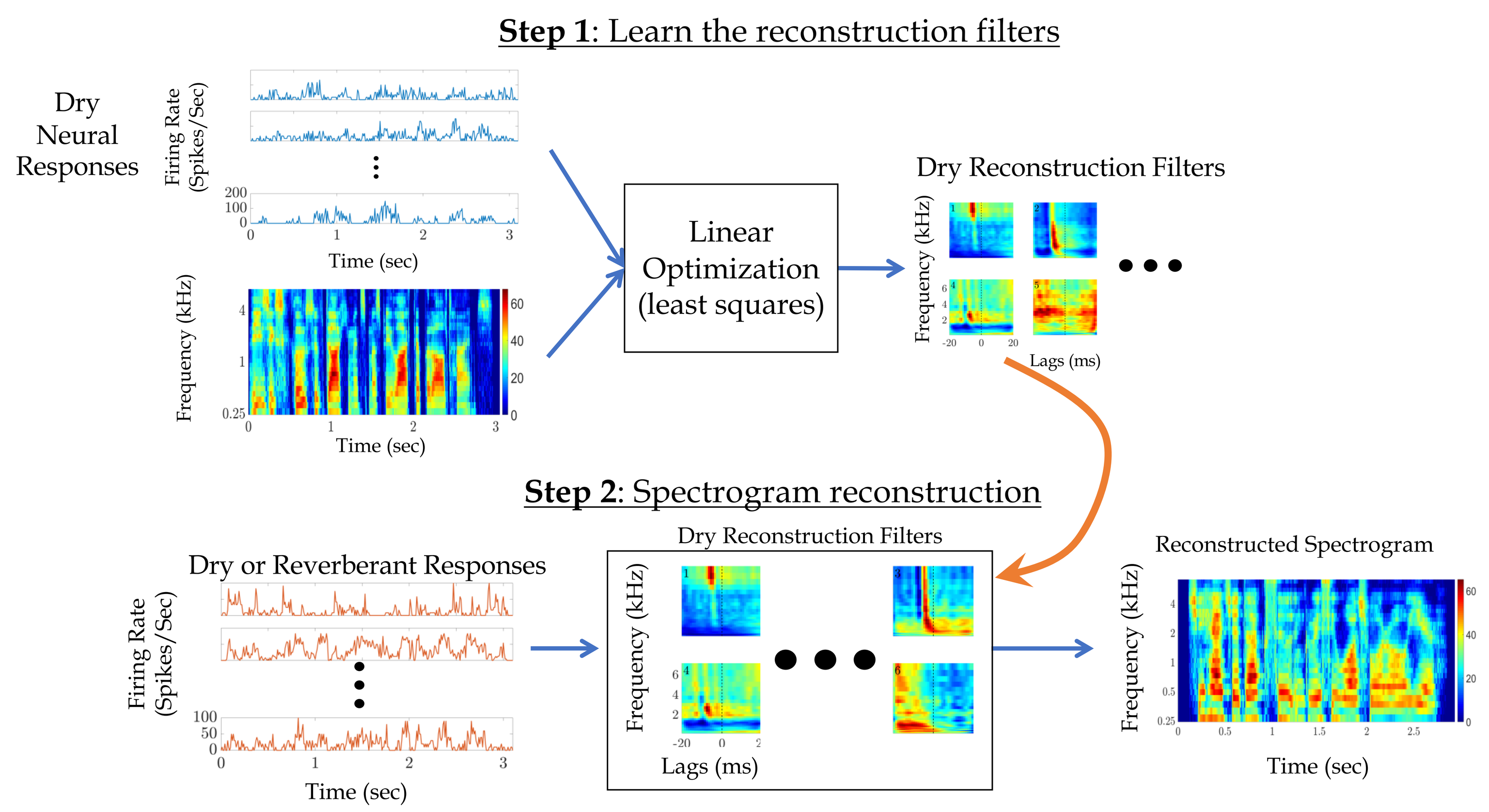
Single unit (SU) and Multi-unit Activity (MUA)

- **Electrophysiology**: To further characterize how reverberant stimuli are processed along the auditory pathway, we recorded single-unit (SU) and multiunit activities (MUAs) from the inferior colliculus (IC) of unanesthetized rabbits in response to speech utterances[6].
- we record single units (SUs) and multi-unit activity (MUA) using linear microelectrode array (LMA, 6-8 recording sites).



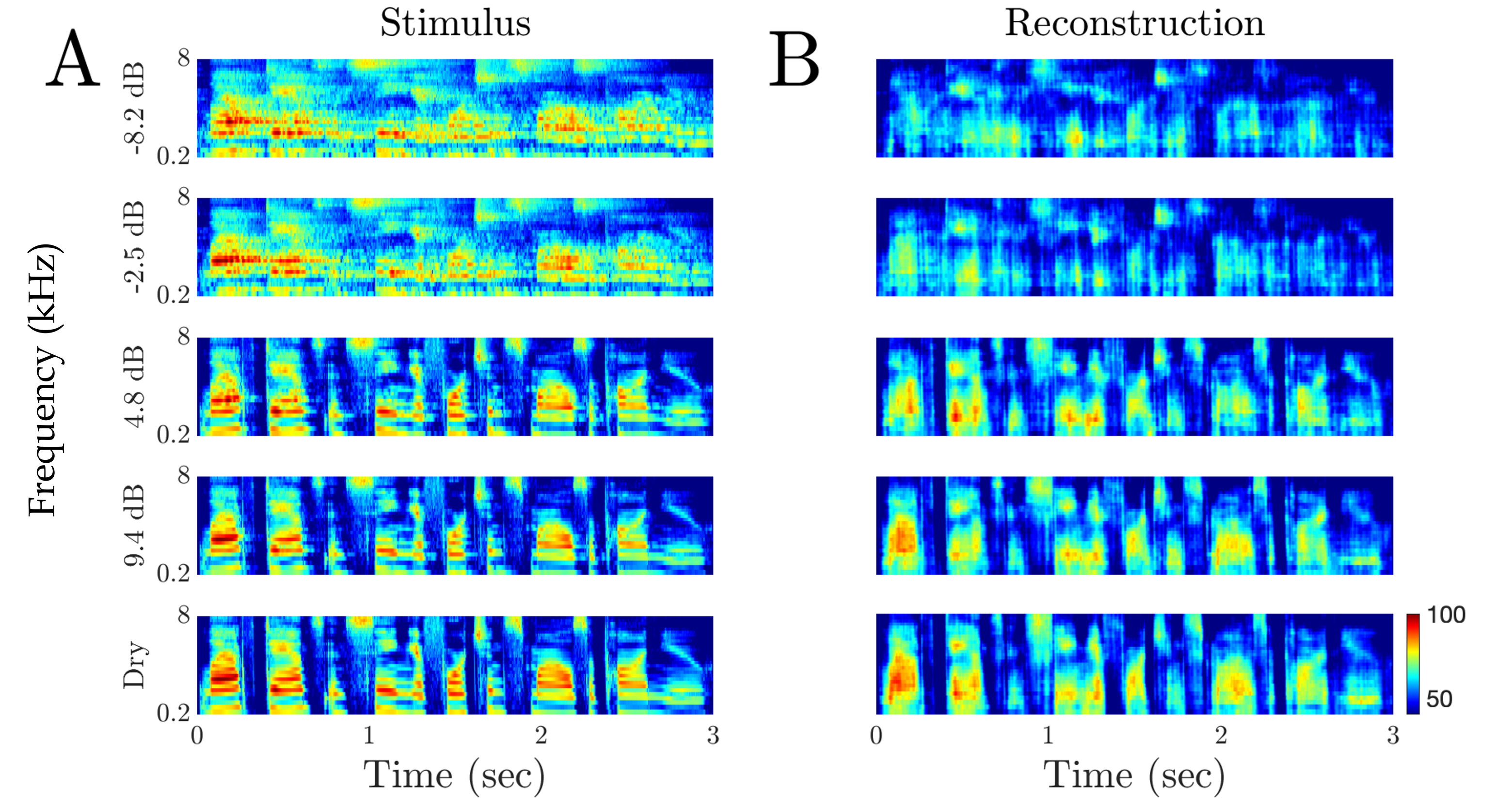
Linear Spectrogram Reconstruction

To quantify the amount of speech information available in the responses of the neural population, we used linear stimulus reconstruction techniques[3]. Optimal mean-squared error reconstruction filters ("dry-filters") were derived from a training set of responses to the dry stimulus condition and for various ensembles of units. We then applied these dry-filters to the responses of both dry (testing set) and reverberant stimuli to obtain reconstructed spectrograms and compared the reconstructed spectrograms to the spectrogram of the original dry speech.



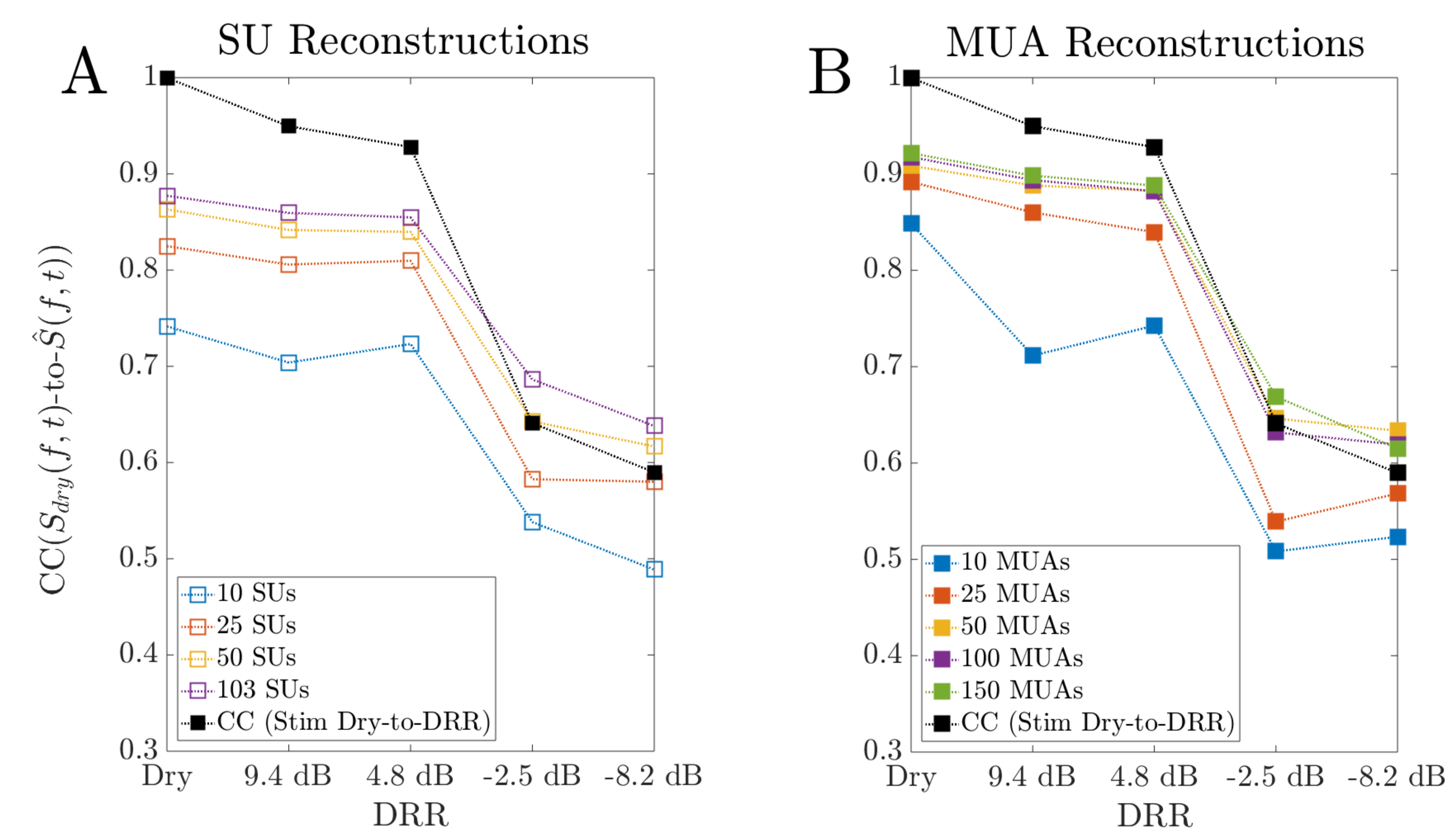
Results

Linear Spectrogram Reconstructions



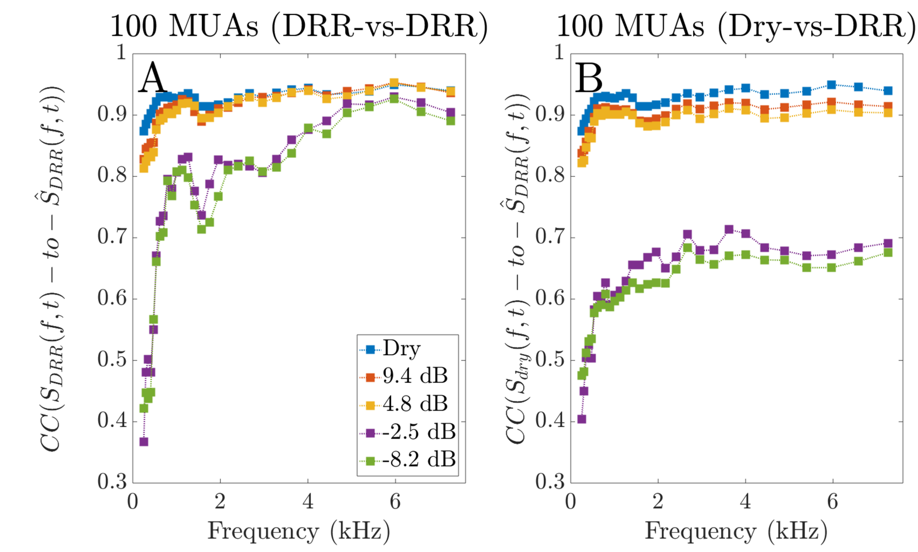
- A female speaker[6] says "Laugh, dance, and sing if fortune smiles upon you".
- On the left (A), stimuli spectrograms of dry and other DRR conditions. On the right (B), the correspondent linear spectrogram reconstructions.
- All reconstructions are performed with dry reconstruction filters.
- 241 MUAs are used for the reconstructions in B.
- Spectrograms has 30 frequency bands (from 250 Hz to 8k Hz); frequencies are distributed on a log scale; step size along the time domain is 5 sec.

Comparing Single Units with Multiunit Activities

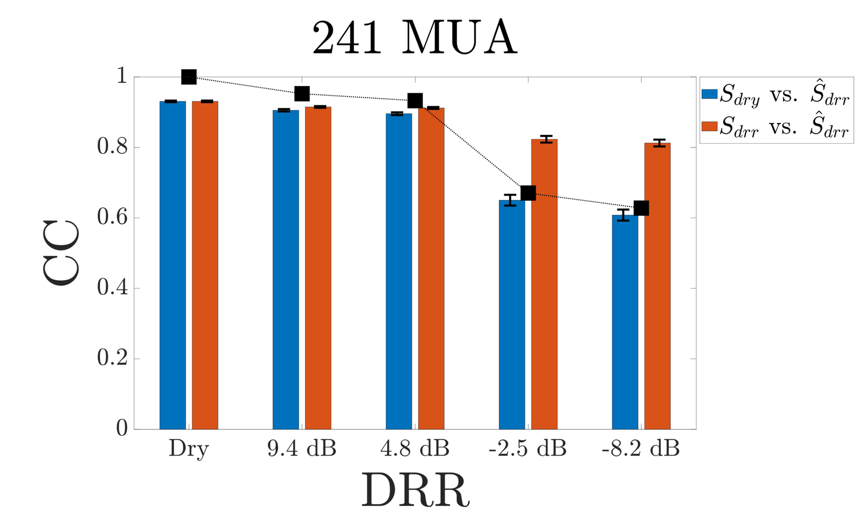


- Spectrograms reconstructed from responses of 25-50 units to dry speech usually showed good resemblance to the original speech spectrograms.
- Reconstructions based on MUAs were more accurate than reconstructions from SU activities (Average Pearson correlation coefficients (CCs) for reconstructions based on 100 units: 0.917 for MUA vs. 0.877 for SU), but for both MUAs and SUs the reconstruction quality deteriorated with increasing reverberation.
- Stimulus CCs (black squares) are computed between stimuli (without responses) and are served as benchmark.

Reconstruction as a Function of DRR Condition



- Spectrogram reconstructions for dry and mild reverberation stimulus condition (DRR > 0 dB) are relatively good (CC > 0.8) but deteriorates in severe reverberation conditions (DRR < 0 dB).
- Spectrograms reconstructed with dry-filters from responses to reverberant stimuli resembled spectrograms of reverberant speech better than spectrograms of dry speech.



- For each reverberant condition, the amount of degradation in the reconstructed spectrogram was proportional to the mean deviation between the reverberant speech spectrogram and the dry spectrogram, i.e., the neural degradation paralleled the degradation in the stimulus.

Conclusions

- Reverberation degrades the responses to speech in the IC of unanesthetized rabbits.
- Linear spectrogram reconstructions from small samples of IC units (25-to-50 units) were good for the dry stimulus condition and robust in mild reverberation (DRR > 0 dB) but deteriorated in severe reverberation conditions (DRR < 0 dB).
- Multiunit activity (MUA) produced better stimulus reconstruction than single unit (SU) responses
- Overall, the results provide no evidence for compensation for the effects of reverberation in neural responses from the rabbit IC when studied with linear reconstruction techniques.

Bibliography

1. Slama and Delgutte, 2015, JNeurosci. 35(10):4452
2. Mesgarani et al., 2014, PNAS 111:6792
3. Mesgarani et al., 2009, JNeurophysiol 102:3329
4. Day, et al., 2013, JNeurosci, 33(40):15837
5. Allen, et al., 1979, JASA, 65(4):943
6. Garofolo, J.S., 1993, *Linguistic Data Consortium, 1993*

Acknowledge

O.B. thank wish to thanks to Ken Hancock for software support, Ishmael Stefanov-Wagner for engineering expertise, and Joseph Wagner for surgical assistance. Supported by NIH Grant R01 DC 002258.