Human cortical auditory sensitivity adaptation to simulated hearing loss: A data-driven fMRI approach with naturalistic speech

Arkan Al-Zubaidi (arkan.al-zubaidi@uni-oldenburg.de)
Applied Neurocognitive Psychology Lab and Cluster of Excellence Hearing4all, University of Oldenburg
Oldenburg, 26129 Niedersachsen Germany

Jochem Rieger (jochem.rieger@uni-oldenburg.de)
Applied Neurocognitive Psychology Lab and Cluster of Excellence Hearing4all, University of Oldenburg
Oldenburg, 26129 Niedersachsen Germany

Abstract:

The naturalistic auditory stimulus reveals dynamic changes in cortical temporal lobe activity as participants engage with continuous speech (Boos et al., 2021). However, there is little knowledge regarding cortical auditory speech processing in response to clear and degraded naturalistic stimuli. In this fMRI study, we used voxel-wise spectrotemporal receptive field (VWSTRF) estimation to assess the neuronal population sensitivity to acoustic features (mel spectrogram [MS]) in different auditory areas for clear stimuli and at two levels of simulated hearing loss (SHL). The audio description of the movie "Forrest Gump" soundtrack was employed as a clear naturalistic stimulus. VWSTRF models with MS features predicted the BOLD responses in the Heschl gyrus and planum temporale. We find that the spectrotemporal sensitivities change in a non-trivial way. The neuronal population sensitivities increased with SHL in low-frequencies (100-1000 Hz), where stimulus degradation was milder, and decreased in high frequencies (4-8 kHz), where stimulus degradation was heaver. This suggests a compensatory sensitivity increase in the frequency range below 1kHz, which is relevant for speech comprehension. Conversely, the observed sensitivity reduction in the high-frequency range (>4 kHz) aligns with the assumption that the neuronal population implements passive filtering in this frequency range, which is less relevant for speech comprehension. In conclusion, the present study demonstrates that the spectrotemporal sensitivities of neural ensembles exhibited rapid, potentially task-relevant changes while participants listened to continuous speech in natural environments with different levels of SHL.

Keywords: simulated hearing loss; naturalistic speech; voxel-wise spectrotemporal receptive field model

Methods

We collected fMRI data from 30 healthy participants with normal hearing who listened to the German audio description of the movie "Forrest Gump" (Hanke et al., 2014). The audio movie was divided into eight segments, each with a different listening condition: Clear stimulus (CS), mild degradation (S2), and heavy degradation (N4), according to Bisgaard et al. (2010). Precisely, in S2, the frequencies below 1kHz are only moderately attenuated by -30 dB, and then there is a steep fall-off at higher frequencies up to -95 dB for simulated mild hearing loss. In N4, the low frequencies are attenuated by -55 dB, and then there is a flat fall-off at higher frequencies (>4 kHz) up to -80 dB for simulated severe hearing loss. VWSTRF models were estimated for each participant and condition separately. To test the generalization of the models, we used a leave-one-segment-out cross-validation schema. We calculated the Pearson correlation coefficients between predicted and observed BOLD response in each left-out segment to evaluate the model quality for each voxel in the temporal lobe. To characterize how SHL impacts the neuronal encoding, the MS weights were extracted from each VWSTRF model and averaged within four regions of interest (ROIs) in the temporal lobe: Heschl gyrus (HS) and planum temporale (PT) and superior temporal gyrus posterior (STGP) and anterior (STGA).

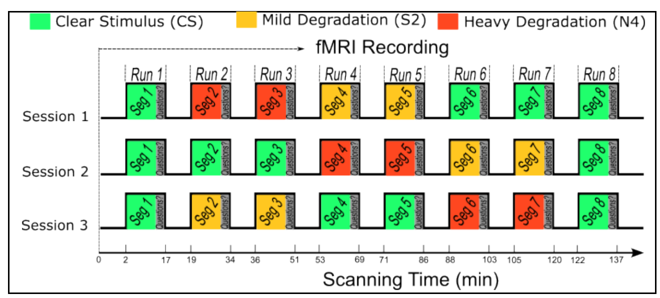

Figure 1: Experimental design. For each participant, the fMRI stimulation protocol consisted of three sessions, each consisting of eight runs. During each session, participants listened to the full movie, but in eight runs, i.e., eight segments were presented in chronological order but at different degradation levels. In three sessions, each segment from 2 to 7 was presented at all degradation levels.

Results

At the perceptual level, speech intelligibility was better in CS than in the S2 and N4 conditions. In contrast, the encoding models best predicted the BOLD responses in N4 compared to the CS and S2 conditions. In the N4 condition, we found the highest correlations between predicted and observed BOLD responses concentrated in the early auditory areas on superior temporal cortex (Heschl's gyrus and planum temporale). The spatial pattern of correlations was less focal in the CS and S2 conditions (Figure 2).

Figure 3 shows the average VWSTRF-derived spectrotemporal sensitivities of different ROIs (across all participants). For instance, in the heavy degradation stimulus (N4 condition), HG showed the highest neuronal population sensitivities for the low-frequencies (100-1000 Hz, orange rectangles) at lagging time 4s compared to the S2 and CS conditions. This neural characterization was the opposite for the high-frequencies (4-8 KHz, black rectangles) in response to the natural stimuli without SHL.

Figure 4 shows the difference between spectrotemporal sensitivities according to the SHL: CS-S2 (green) and CS-N4 (blue). For instance, in early auditory areas (Heschl's gyrus and planum temporale), compared to the CS, the spectrotemporal sensitivities of heavily degraded stimuli (N4) were higher at low frequencies(100-1000 Hz) and were lower at high frequencies (4-8 kHz).

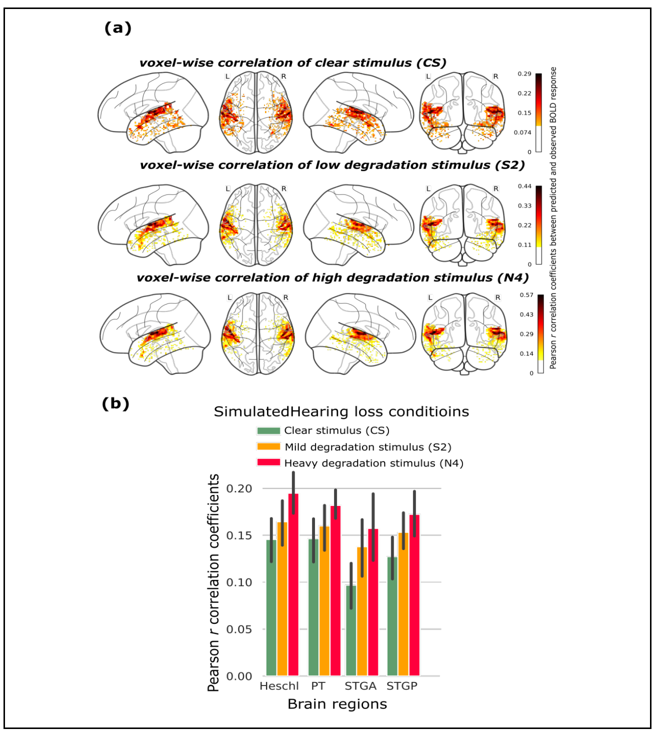

Figure 2: Pearson correlation coefficients between predicted and observed BOLD response for each voxel (a) an exemplary participant and (b) different brain regions in the temporal lobe averaged across participants.

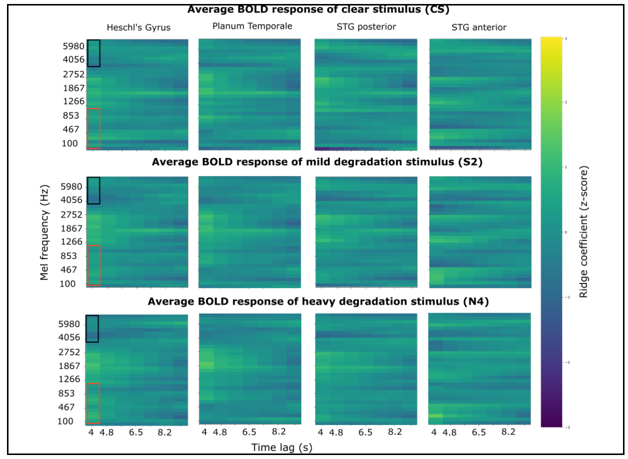

Figure 3: Average VWSTRF-derived spectrotemporal sensitivities of four temporal lobe regions at three speech levels. All spectrotemporal sensitivities shown have a color scale in z-score units.

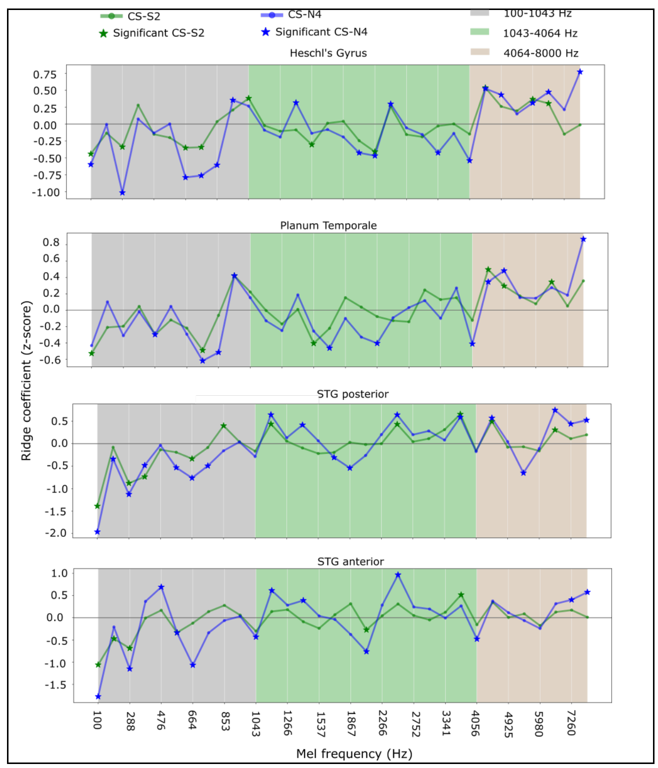

Figure 4: Difference between spectrotemporal sensitivities at time lag 4s. The sensitivity increased with the SHL in low-frequency bands (100-1000 Hz), indicating a minimal level of stimulus degradation. Conversely, a decrease in sensitivity was observed across high-frequency bands (4-8 kHz), reflecting a more pronounced level of degradation.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/1 - Project ID 390895286.

References

Bisgaard, N., Vlaming, M. S., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in amplification*, *14*(2), 113-120.

Boos, M., Lücke, J., & Rieger, J. W. (2021). Generalizable dimensions of human cortical auditory processing of speech in natural soundscapes: A data-driven ultra high field fMRI approach. *NeuroImage*, *237*, 118106.

Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., ... & Stadler, J. (2014). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Scientific data*, *1*(1), 1-18.