# Supplementary Material to the MICCAI 2024 paper entitled "MEGFormer: enhancing speech decoding from brain activity through extended semantic representations"

## 1 Closeness of phonemes audio and MEG segments

The Fig. 1 represents the histograms described in the main paper in the Results section for MEG channels over the auditory cortex. A normalised version of the DTW metric we used was computed as follows:

$$nDTW(audio,\ MEG) = 1 - \frac{DTW(audio,\ MEG)}{\max\{N_{audio}, N_{MEG}\}} \tag{1}$$

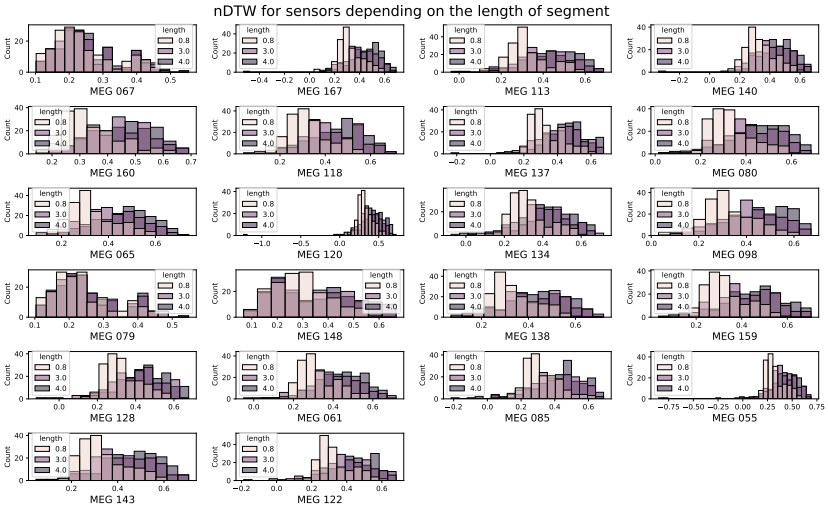


**Fig. 1.** Histograms of nDTW (eq. 1) values for subjects of Gwilliams dataset. nDTW was evaluated for 3 lengths of MEG signals $(0.8, 3, 4s)$ represented in legends.

Fig. 2 contains the Mean Square Error (MSE) between Short-Time Fourier Transform (STFT) of MEG sensors over the auditory cortex. Each value is the sum of MSE between STFTs of MEG signals from corresponding sensors. This sum was evaluated for all phonemes included in one session. Fig. 3 show the same MSE for $\alpha, \beta, \gamma$ frequencies ranges. The $\alpha$ and $\beta$ frequencies show more

contrast in the distance than the $\gamma$ range. We consider that it's only the feature of the MEG system to analyse these ranges more sensitively.

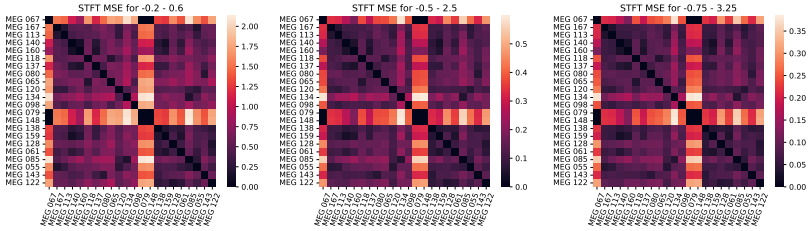


**Fig. 2.** Several clusters of close sensors (dark color) become closer with the MEG segment's length increasing. Greater distances (brighter colors) are also reduced showing a tendency to join the clusters.

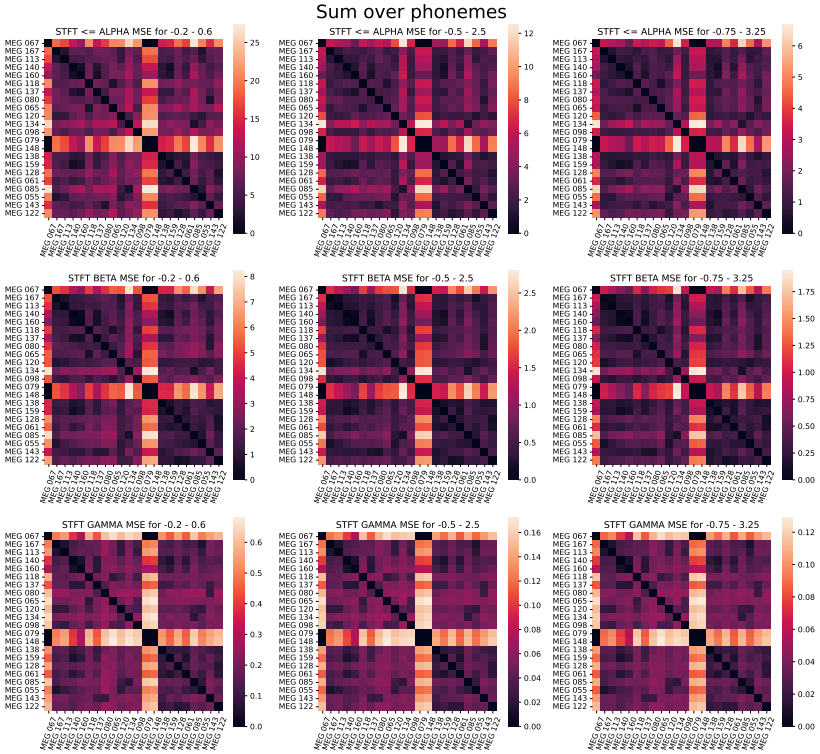


**Fig. 3.** MSE between STFTs for frequencies bands of sensors over the audial cortex.