A Appendix

A.1 Squeeze and excitation network

We adapt the SE block from work [6] to process 1D data and only use it for the first layer. The SE block mainly consists of two parts: (1) a squeeze net $F_{sq}(\cdot)$, which in our case is a adaptive average pooling which reduces T timesteps to 1 with a learnable kernel. (2) an excitation net $F_{ex}(\cdot, W)$, in our case a fully connected network which can be represented as $F_{ex}(\cdot, W) = Sigmoid((W_2(ReLU(W_1(X')))))$ where $W_1 \in \mathbb{R}^{C|16 \times C}, W_2 \in \mathbb{R}^{C \times C|16}$. It takes the output of the squeeze net X' as input and generates channel-wise scaling factors $s \in \mathbb{R}^{1 \times C}$. These scaling factors are then multiplied channel-wise with the original input X. Figure 5 shows an illustration of the adapted SE block we use in this work.

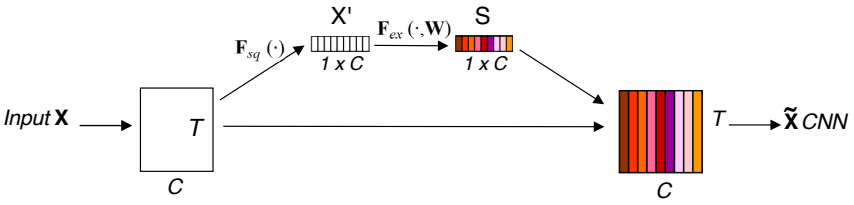

Figure 5: SE block on the input layer to determine the scaling factors of the channels. This figure is adapted from [6].

A.2 Random electrode selection

Some electrodes were implanted further away from the location where the stimulus was placed. We filter out these electrodes before doing a random selection for the random electrode selection experiments in Figure. 4.

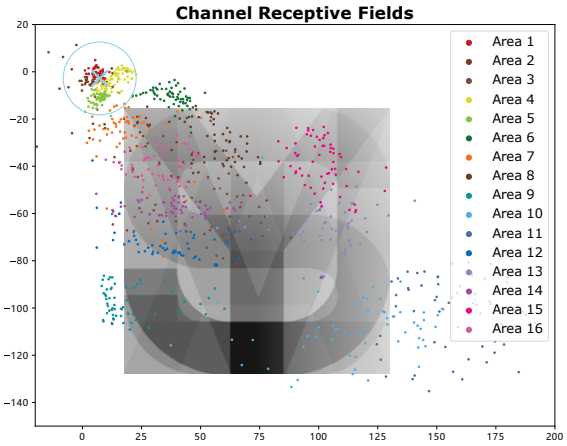

Figure 6: An example electrode whose receptive field (RF) does not overlap with the position where the visual stimuli is placed. The light blue cross marks the RF center of a specific electrode and the light blue circle marks the RF extent. Figure is adapted from [4].

A.3 Choosing important electrodes

Figure 7 shows the top-k important electrodes determined by ranking the channel importance weights of the SE-module used in Figure 4.

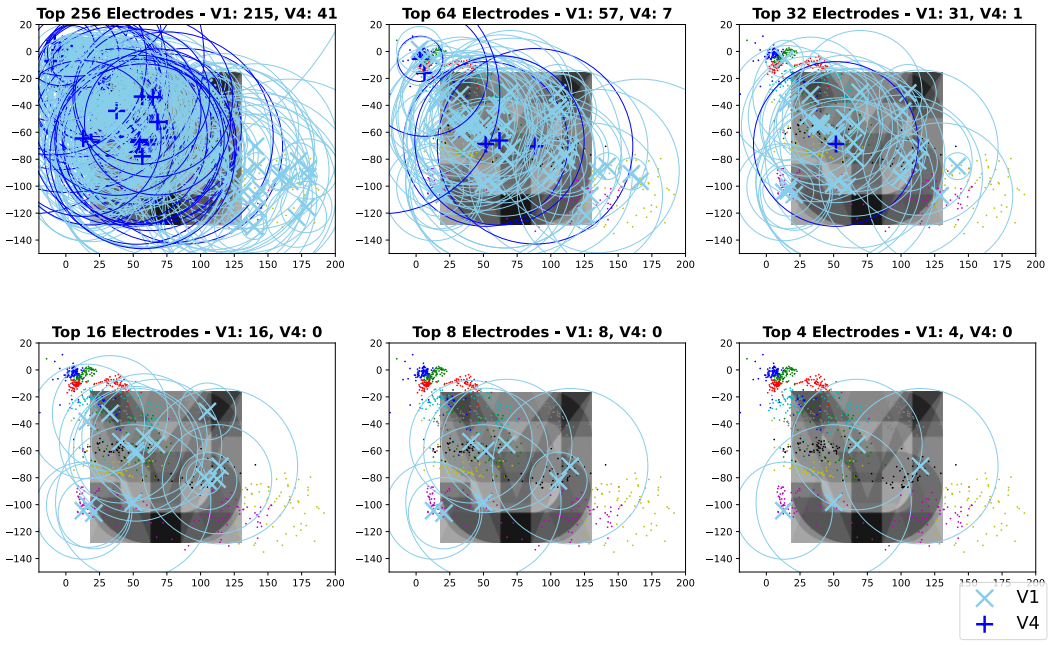

Figure 7: The receptive fields of the top-k most important electrodes corresponding to those with the largest SE scaling factors.