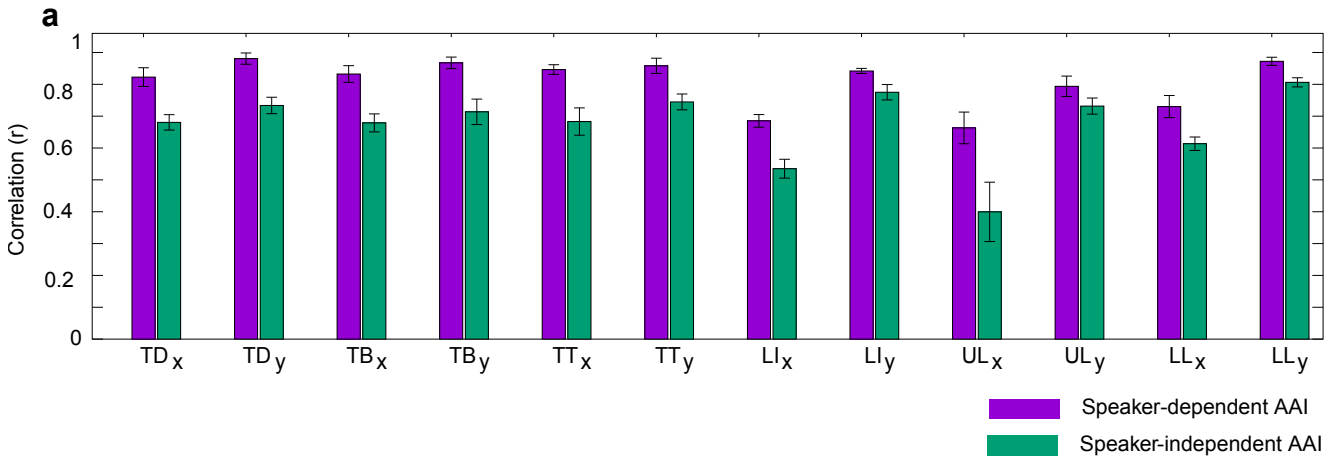
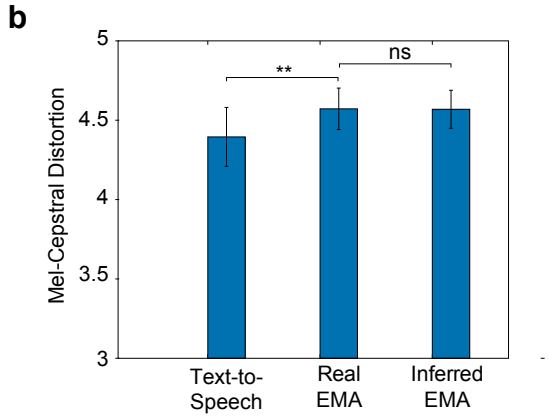
Supplementary Information





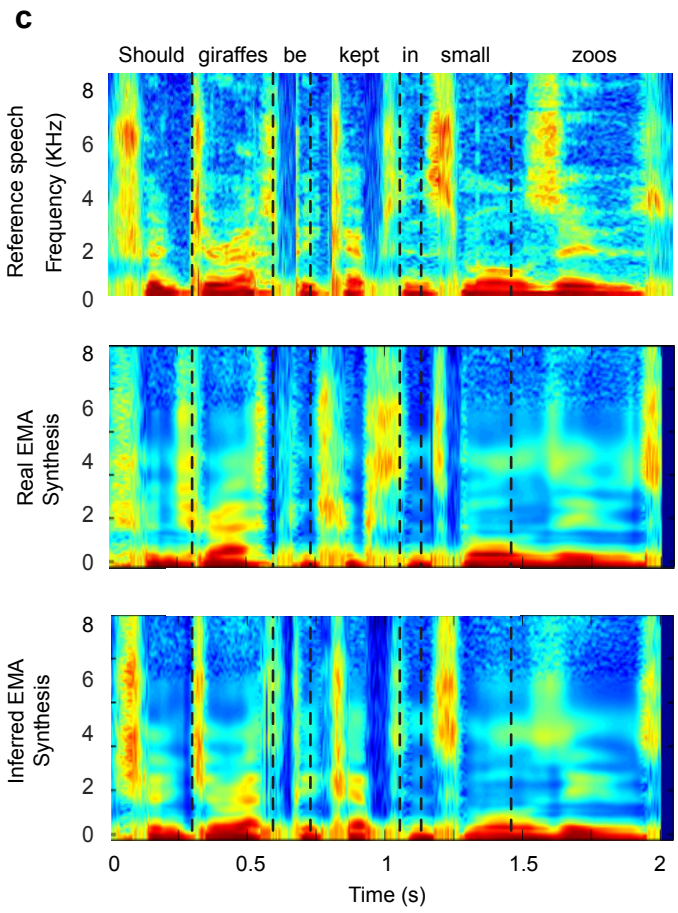


Figure S1. Acoustic-to-Articulatory Inversion, related to Figure 1. a, Articulator-wise breakdown of performance (pearson's correlation coefficient) of inferred articulatory trajectories. Speaker-dependent AAI refers to AAI using articulatory data from the speaker to train the model while speaker-independent AAI refers to AAI that does not use any articulatory data from the speaker to train the model (used in present study). **b**, Performance of articulatory synthesis in terms of mel-cepstral distortion comparing EMA based synthesis for both real and inferred EMA. General text-to-speech synthesis (phonemes, context, duration) shown for reference. **c**, Comparison of an example synthesized utterance based on real EMA and inferred EMA. Top most spectrogram is of the speaker's actual production for the utterance.

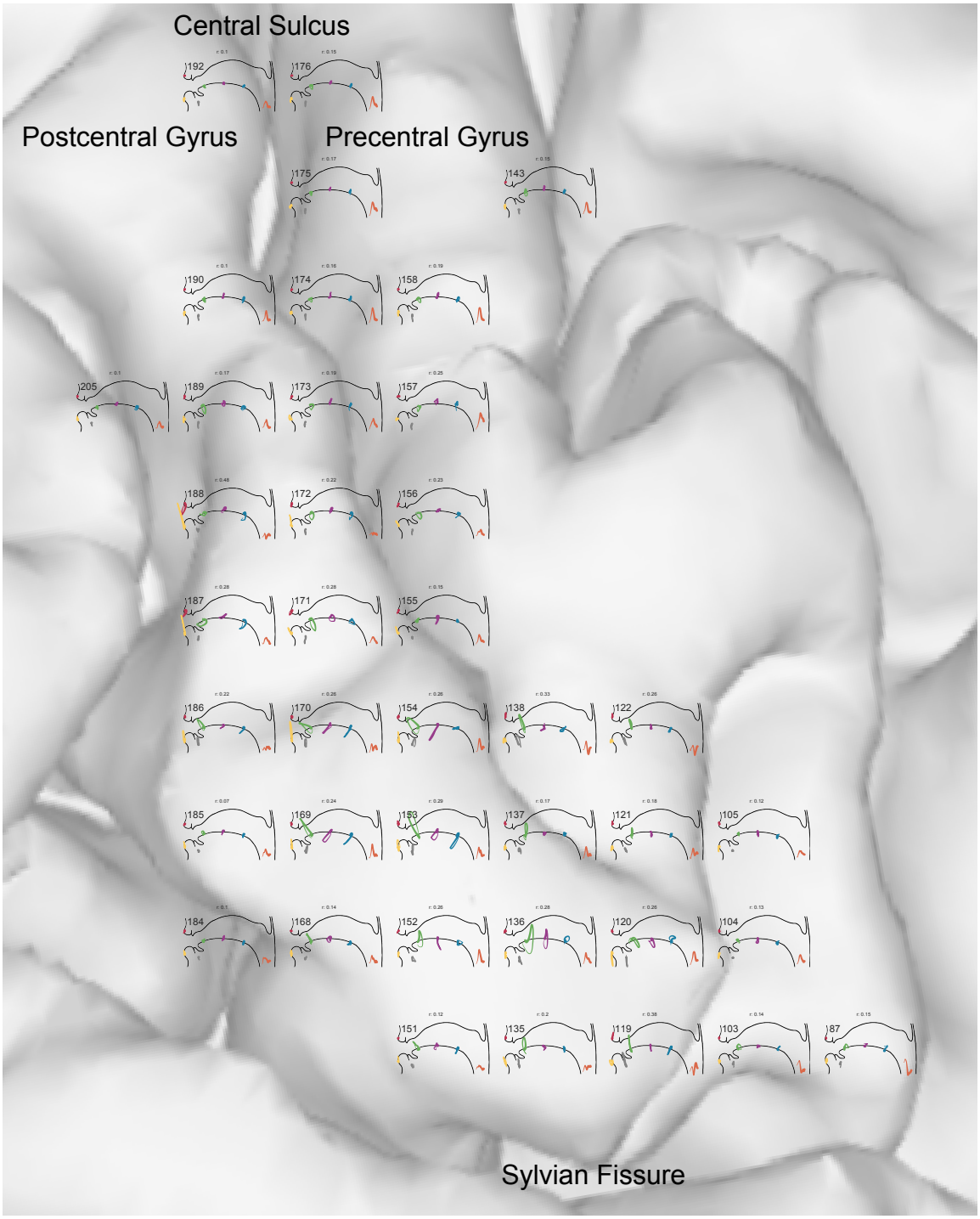


Figure S2. Articulatory kinematic trajectories for an individual subject, related to Figure 2. For one subject with a right hemisphere ECoG grid, we plotted the encoded articulatory kinematic trajectories (AKTs) for each electrode in correspondence to its location on the cortical surface. Electrodes not active during speech were not shown. Each vocal tract plot shows the encoding model filter weights. Thin to thick lines indicate the time course of each articulator trajectory. We found AKTs were encoded in both the precentral and postcentral gyri. Furthermore, AKTs with similar trajectory shapes appeared to spatially close to one another. r values for the correlation of each encoding model with high gamma are shown above the vocal tract plots.

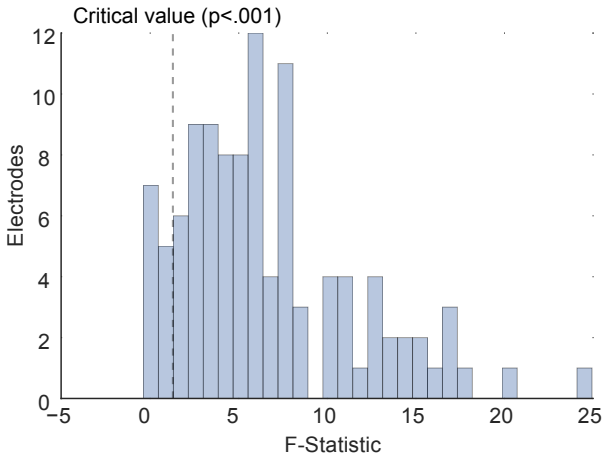


Figure S3. Electrodes encode coordinated articulatory movements involving multiple articulators, related to Quantification and Statistical Analyses. We used a nested regression model to compare whether the additional variance explained by trajectories of multiple articulatory was significant when compared to the variance explained by the trajectory of a single articulator. Since differences in parameter numbers can influence the explained variance solely by changing model complexity, we computed an F statistic on held-out data for each electrode to statistically test for model significance. Here, we plotted the distribution of F statistics and found that 96 out of 108 electrodes had F statistic greater than the critical value ($F(280, 1820)>1.31$, $p<.001$). The mean F statistic was 6.68 indicating that single electrodes encoded coordinated trajectories of multiple articulators.

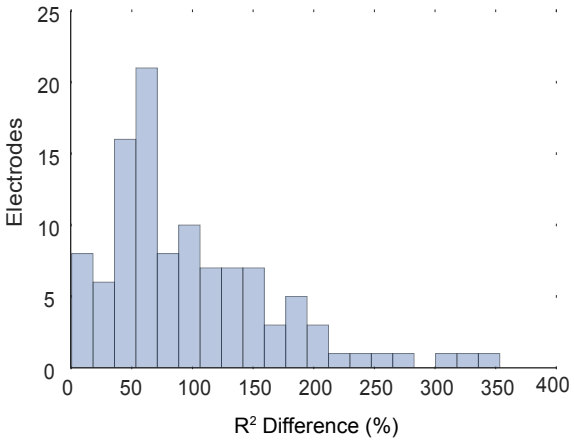


Figure S4. Comparison of variance explained by AKT model over single articulator model, related to Quantification and Statistical Analyses. The change in explained variance on held-out data by using the AKT model (all articulators) instead of the single articulatory trajectory model (X and Y for one articulator) is shown for each electrode as a percentage. The mean increase in explained variance from the AKT model was 99.55% +/- 8.63%.

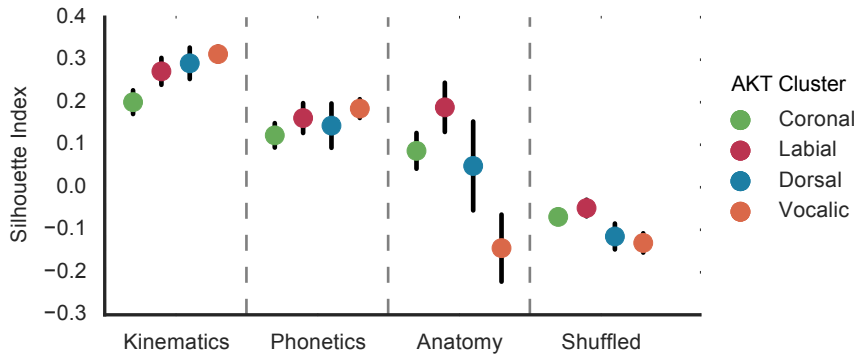


Figure S5. Silhouette analysis of clusters, related to Figures 3 and 4. We quantified the relationship between within-cluster and between-cluster similarities for each AKT cluster using the silhouetteindex as a measure of clustering strength. A silhouette index close to 1 indicates that the electrode is highly matched to its own cluster. 0 indicates that that the clusters may be overlapping, while -1 indicates that the electrode may be assigned to the wrong cluster. The degrees of clustering strength of AKT clusters for kinematic and phonetic descriptions were statistically significant above clusters of shuffled AKTs indicating that clusters had both similar kinematic and phonetic outcomes ($p < .01$, Wilcoxon signed rank tests). However, only the spatial clusterings for coronal and labial AKTs were statistically significant ($p < .01$, Wilcoxon signed rank tests). Only one subject had more than two dorsal AKT electrodes and we could not justly quantify the clustering strength of this cluster. The low silhouette index for anatomical clustering of the vocalic AKTs was expected because two spatial clusters were later found.

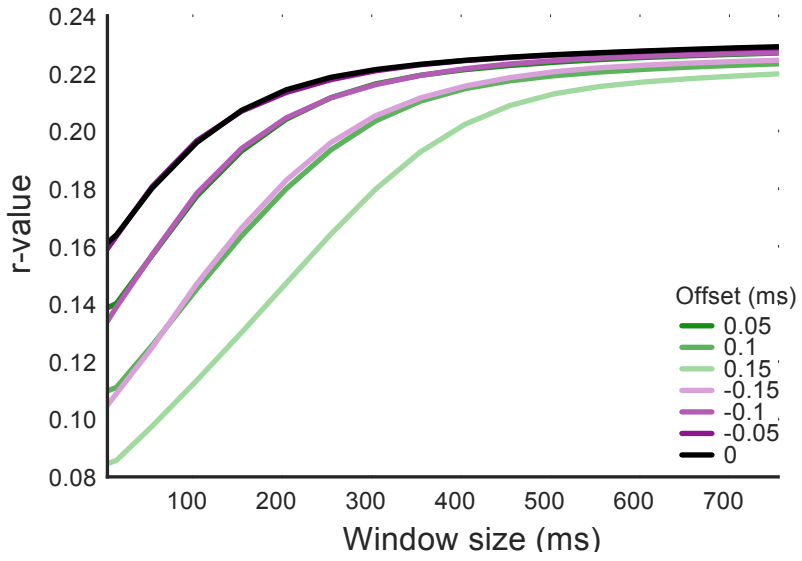


Figure S6. Effect of window size and offset from neural activity on AKT model performance, related to Method Details. Using varying window sizes of articulatory movements and offsets of those movements to electrode activity, the AKT model was fit and then tested on data held-out from training to compute mean correlations (r-value) between high gamma and predicted high gamma activity across all electrodes from all subjects in the study.