Complementary Neural Encoding of Invisible Articulatory Synergies

Summary

During speech perception, endogenous neural oscillations are phase-aligned to the physical regularities of the incoming speech stimulus, a phenomenon known as "speech entrainment". This synchronization facilitates the tracking of linguistic units and aids speech comprehension. However, pure bottom-up encoding of the acoustical input falls short in explaining phenomena where perceptual experiences deviate from the true acoustical content of the scenario (e.g., phonemic restoration and the cocktail party effect). Crucially, speech signals result from brain orchestration of forward and backward acoustic-articulatory mapping. During speech perception, similar processes are thought to produce top-down signals to select a relevant low-dimensional manifold from the input space. Here we explore whether, during speech listening, the brain simulates the articulatory movements of the speaker and investigate its role in tasks of variable difficulty. Participants listened to sentences and performed a speech-rhyming task. Real kinematic data of the speaker's vocal tract, recorded via electromagnetic articulography (EMA), were aligned with corresponding acoustic outputs. Articulatory synergies were extracted from the EMA data using Principal Component Analysis (PCA). We then employed Partial Information Decomposition (PID) to segregate the electroencephalographic (EEG) encoding of acoustic and articulatory features into unique, redundant, and synergistic atoms of information. Remarkably, our results reveal that the brain retrieves unique and highly specific information from tongue movements, inaccessible through vision. We then median-split sentences into easy (ES) and hard (HS) based on participants' performance and found increased encoding of unique articulatory information in the theta band under greater task difficulty. We conclude that fine-grained articulatory reconstruction complements the encoding of speech acoustics, underscoring the role of *motor processes* in supporting speech perception.

Materials and Methods

The study employed stimuli from a dataset providing aligned audio and articulatory data using electromagnetic articulography (EMA). Each session involved one participant deciding if a presented word rhymed with a previously heard sentence. EEG data from a 64-channel system were acquired during the whole session and then pre-processed. Responses to the rhyming task were pooled across subjects to classify sentences as easy (ES) or hard (HS) based on subjects' performance. Acoustic stimuli underwent speech envelope extraction (SE), while kinematic signals were dimensionally reduced using Principal Component Analysis (PCA), retaining the first four PCs (PCi, i=1,…,4). Acoustic and kinematic features are strongly related to each other, as speech acoustic outputs are directly constrained by phonoarticulatory movements and could therefore produce overlapping information. We used Partial Information Decomposition (PID) (Williams and Beer, 2010; Ince, 2017) to compute the information encoding between two sources (SE and each PCi) and one target (each EEG channel) to extract four atoms of partial information: a) information exclusively provided by each of the two sources (Unq(SE) and Unq(PC)); b) information only available when the two sources are together encoded (Syn(SE,PC1)); c) information shared between the two sources (Rdn(SE,PC1)). We used one-tailed cluster-based permutation statistics applied for group-level testing, comparing original information values against a surrogate distribution of circularly shifted data. The PID analysis was repeated for ES and HS separately, in both delta- and theta-bands, and tested the hypothesis that multisensor information values for ES and HS belonged to the same distribution.

Results and Conclusions

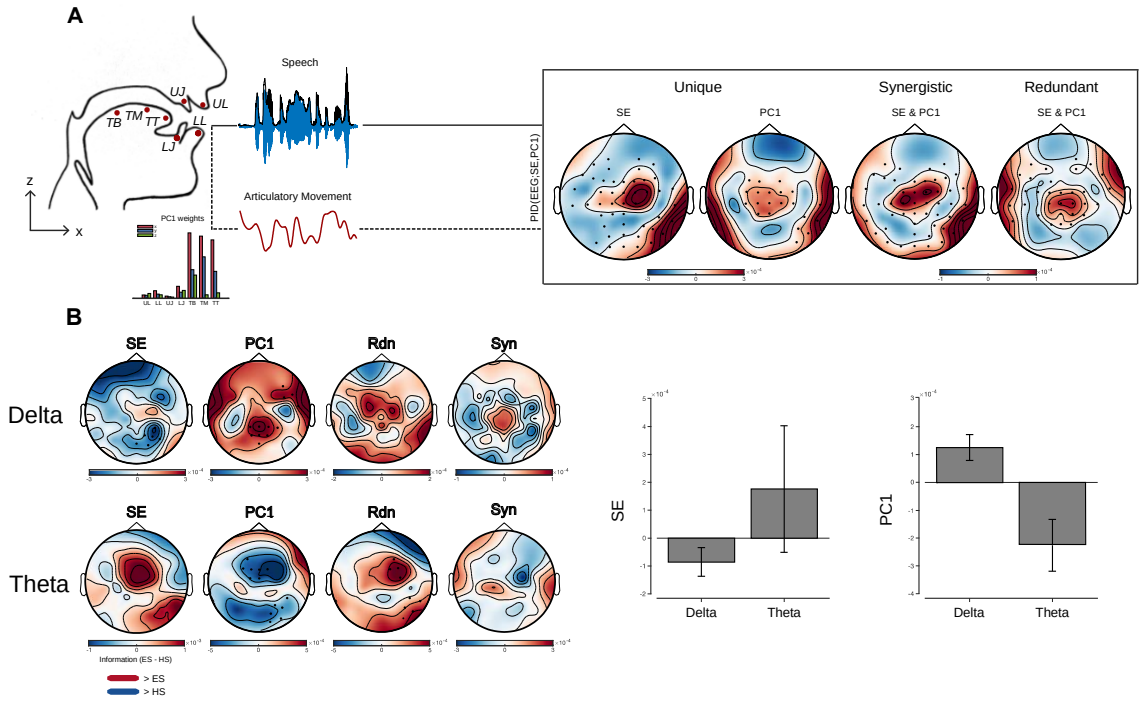

Figure. PID results. A. Schematic of features extraction (left); topographical distribution of across-subjects average information values obtained by PID analyzes performed on the broad-band filtered data (0.5–10 Hz) (right). B. Topographical distributions show the mean of the information difference across subjects (ES- HS) for the unique, redundant, and synergistic atoms of information obtained for band-pass filtered data in the delta and theta bands (left); group average and standard error of the mean (SEM) of the difference between the average information across all channels in the two conditions (ES- HS). Black dots highlight the electrodes belonging to the clusters that survived the test statistics (alpha level = 0.05).

Speech perception requires the brain to integrate diverse information sources based on reliability. Indeed, when the acoustic signal is compromised, increased reliance on visual cues occurs. This study shows that speech processing relies on motor signals for precise articulatory reconstruction, that, following the rule of inverse effectiveness, complements acoustic encoding during more challenging listening tasks. Most importantly, the encoding of tongue movements, inaccessible through vision, cannot result from passive exposure to environmental statistics. Instead, it requires a speech-producing agent that leverages learned acoustic-articulatory mappings. Altogether these findings underscore the role of *motor processes* in speech perception (Pastore et al., 2022).

Bibliography

Williams PL, Beer RD. Nonnegative decomposition of multivariate information. https://doi.org/10.48550/arXiv.1004.2515.

Ince RAA. Measuring Multivariate Redundant Information with Pointwise Common Change in Surprisal. Entropy 19, 2017. doi: 10.3390/e19070318.

Pastore A, Tomassini A, Delis I, Dolfini E, Fadiga L, D'Ausilio A. Speech listening entails neural encoding of invisible articulatory features. Neuroimage 264: 119724, 2022. doi: https://doi.org/10.1016/j.neuroimage.2022.119724.