

Seeing a speaker's face enhances speech intelligibility in adverse environments. We investigated the underlying network mechanisms by quantifying local speech representations and directed functional connectivity in MEG data obtained while human participants listened to speech of varying acoustic SNR and visual context in two paradigms. One study involved the presentation of long (6min) continuous texts and manipulated SNR across four levels. We found that during high acoustic SNR speech encoding as reflected by temporally entrained brain activity (mostly in delta and theta bands) was strong in temporal and inferior frontal cortex, while during low SNR entrainment emerged in premotor and superior frontal cortex. These changes in local encoding were accompanied by changes in directed connectivity along the ventral stream and the auditory-premotor axis. Importantly, in this study the behavioural benefit arising from seeing the speaker's face was not predicted by changes in local encoding but rather by enhanced functional connectivity between temporal and inferior frontal cortex. A second paradigm tested audio-visual perceptual benefits directly at the single sentence level using a word recognition task. We used single-trial decoding to characterize local representations of word identity and relied on measures of intersection information to directly link these neural representations to single trial perception. This revealed a network of superior temporal, supramarginal and inferior frontal regions, in which the neural read-out of word identity carries predictive power for perception in multisensory conditions. These findings highlight the need to consider both, local representations and functional connectivity, when trying to elucidate the neural underpinnings of speech perception and highlight how both, ventral and dorsal pathways, facilitate speech comprehension in challenging multisensory environments.