Abstract

Primates are social animals whose communication is based on their conspecifics' vocalizations and facial expressions. Although a lot of work to date has studied the unimodal representation of vocal or facial information, little is known about the way the nervous system supports the processing of communication signals from different sensory modalities to combine them into a coherent audiovisual percept. It is thought that the brains of human and nonhuman primates evaluate vocal expressions and facial information separately in specialized 'voice' and 'face' brain regions but we wondered if cross-sensory interactions were already evident at the neuronal level in these typically unimodal brain regions.

Using movies of vocalizing humans and monkeys as stimuli, we recorded extracellularly from the auditory cortex of a macaque monkey, targeting his 'voice' region in the right hemisphere. Within a multi-factorial design we evaluated how these auditory neurons responded to different sensory modalities (auditory or visual) or combinations of modalities (audiovisual). We also analyzed the responses for species-specific effects (human/monkey speaker), call type specificity (coo/grunt), as well as speaker familiarity, size and identity.

Following the approach in the original fMRI study localizing the monkey voice region, our recordings identified a voice area 'cluster' in this animal. Within this auditory cluster of sites, we observed a significant visual influence on both the local-field potential (LFP) and the spiking activity (AMUA), and found that 30% of the sites showed audiovisual interactions in the LFP signals, and 38% in the AMUA. Grunts were especially effective stimuli for this region and rather than a specialization for monkey vocalizations, human vocalizations also elicited strong responses.

Our results provide evidence for visual influences in what has been characterized as an auditory 'voice' area suggesting that at least the 'voice' regions are influenced by the visual modality. Voices and faces seem to already interact at traditionally unisensory brain areas, rather than cross-sensory information being combined only in higher-level, associative or multisensory regions of the brain.