
Abstract:
We perceive the world in several modalities, of which seeing and hearing are the most prominent. We see the world in a kaleidoscope of colours and shapes, but only when there is sufficient light and when the line of vision is not blocked. In contrast, we hear across visual barriers, with no need for light and no need to fix our gaze; but hearing fails when the competing noise overwhelms, or when the distance is too great. Archaic humans presumably communicated with both gestures and vocal sounds at first, with sounds not yet distinguished into speech and song. Vocal sounds could be pleasing in friendship and courting, or threatening in conflict; they could coordinate or synchronise group activities.

Speech began to overtake gestures when certain sounds became conventionally paired with meanings, giving rise to words; conventions quickly followed for forming complex words, phrases, and sentences. Similarly in song, syllables became differentiated by pitch and grouped into rhythmic patterns. Since both speech and song share principles of organisation for sensory-motor and cognitive functions, it is natural that there are many (though not all) neural circuits in the brain that serve them both. It is also natural that speech without melody combines with song without words to produce song with words. Vocal sounds became supplemented or complemented with instrumental sounds of various sorts at least 20 kya; they also led to a surrogate system of symbols in the form of writing some 7 kya.

Given the very close relationship between speech and song, exploring methods for recovering speech in certain language disorders is a promising direction of research for both basic neurolinguistics and clinical applications.