CellPress

Review

Recognizing People in Motion

Galit Yovel[1,*] and Alice J. O'Toole[2]

Natural movements of the face and body, as well as voice, provide converging cues to a person's identity. To date, person recognition has been studied primarily with static images of faces. Face recognition, however, is part of a larger system, whose preeminent goal is to efficiently recognize dynamic familiar people in unconstrained environments. We present a comprehensive framework for understanding person recognition as it happens in the real world. In this framework, dynamic information plays the central role in binding multi-modal information from the face, body, and the voice to achieve robust and highly accurate recognition. The superior temporal sulcus (STS) integrates multisensory, dynamic information from the whole person for recognition, thereby complementing its role in social cognition.

Motion Plays a Central Role in Person Recognition

The utility of **biological motion** (see Glossary) for person recognition is seen most clearly when we consider the problem of recognizing familiar people in viewing conditions commonly experienced in the natural world. We argue that motion acts as the key element for binding together faces, bodies, and voices into a coherent representation of a person that supports recognition. In making this argument, we review recent evidence on the role of the body and biological motion in person recognition and consider studies showing that the integration of moving faces and voices facilitates recognition.

The idea that motion plays a central role in person recognition is not prominent in the cognitive science literature. This is due to three longstanding biases in this literature: the common practice of studying person recognition using static images of faces; the emphasis on studying the role of the body and body motion in action perception and social communication, rather than in recognition; and the strong emphasis on the study of human voices for speech perception, rather than for recognition. These biases are justified in decades past, when there was limited knowledge of the psychological, neural, and computational complexities involved in processing a moving and talking person in an unconstrained environment. Recent years have seen advances in all three disciplines that lead us to reconsider person recognition in this more complex *in situ* form. We will show that this complexity simplifies the problem of recognition by distributing the burden over multisensory identity cues and by using motion to bind these cues together. We will argue that the dynamic and multisensory nature of identity cues that apply in the real world point to the **STS** as the likely integration site for person recognition cues with acoustic and spatiotemporal extent.

Axiomatic to this argument is the observation that the visual and auditory person recognition systems in the brain evolved over hundreds of millions of years to work efficiently in the natural world. Recognizing people from a safe distance (at which the face is not sufficiently resolved to be helpful) is a distinct evolutionary advantage. Accordingly, we begin this discussion where perception begins: at a distance. From this perspective, it becomes clear that person recognition is a process that unfolds over space and time, not a decision made to a snapshot of a visual event (Figure 1A, Key Figure). At a distance, person recognition is strongly constrained by the viewing environment and the limits of human sensory systems. Identity cues from the body and

Trends

We propose a comprehensive framework for understanding real-life person recognition. In this framework, person recognition often begins at a distance, where biological motion, body, and voice cues to identity can be highly reliable.

Person recognition in real life benefits from using multiple sources of information, including the face, body, voice, and biological motion.

Dynamic information, in the form of dynamic identity signatures, plays an important role in binding together the face, body, and voice into a multi-modal dynamic representation of a person.

Dynamic identity signatures from the face, body, and voice are used to determine person identity.

We propose the STS as a neural hub for integrating different sources of motion-based identity information from face, body, and voice. This assigns STS an important role in the processing of both invariant and changeable aspects of the dynamic whole person.

[1]School of Psychological Sciences, Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
[2]School of Behavioral and Brain Science, University of Texas at Dallas, Dallas, TX, USA

*Correspondence: gality@post.tau.ac.il (G. Yovel) and otoole@utdallas.edu (A.J. O'Toole).

CellPress

Key Figure

The Roles of Body, Voice, and Biological Motion in Person Recognition

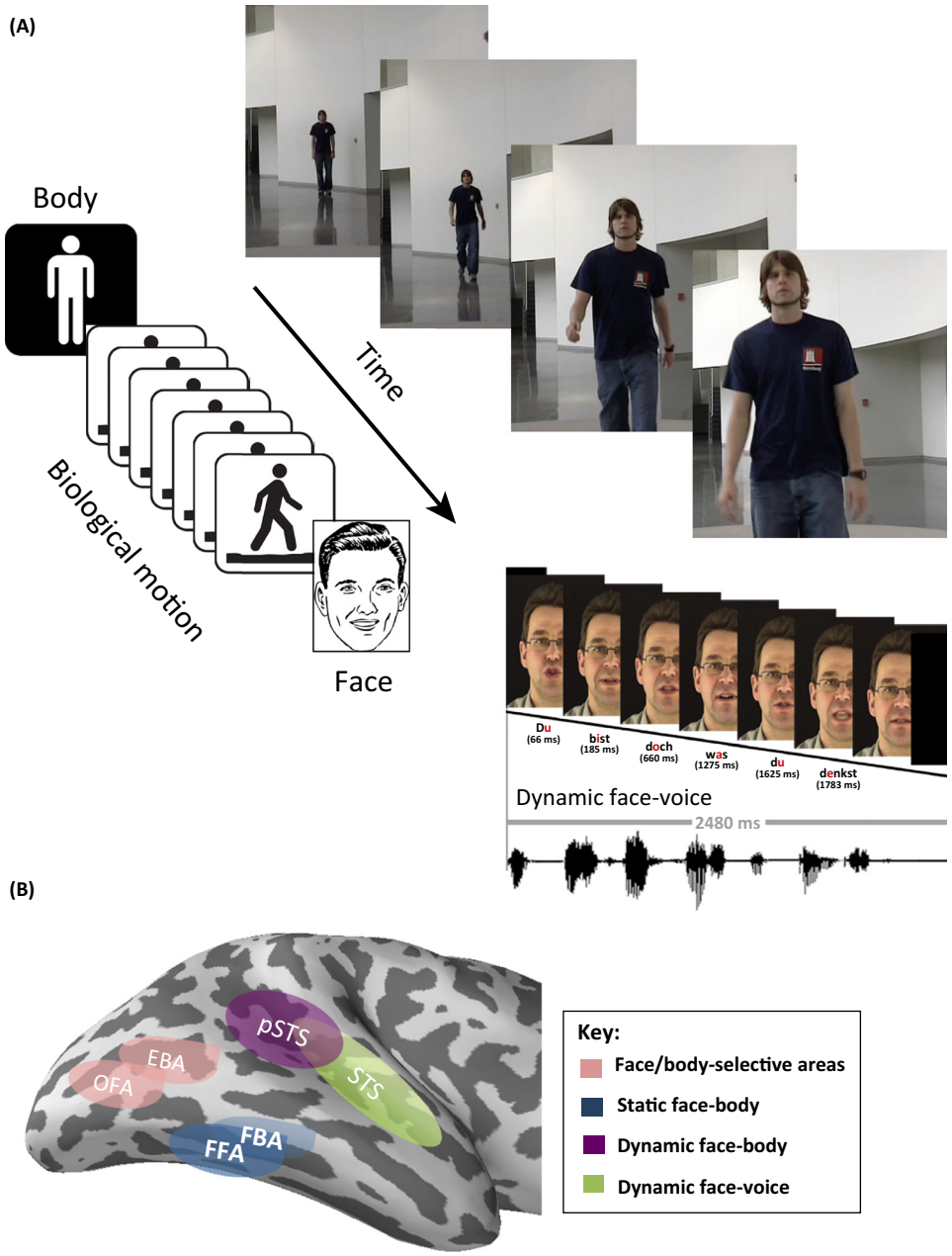

Figure 1. (A) Schematic shows the prominence and primacy of body and biological motion cues to identity at a distance, with the face becoming clearly resolved only at closer distances. This process unfolds over time, allowing for continuous accumulation of identity cues until a confident response is possible. The yoking of a dynamic face and the production of speech is shown below (reproduced, with permission, from [43]). The acoustic and spatiotemporal cues are perceptually bound together by motion. (B) Neural mechanisms of person recognition in the human brain: the posterior face and body areas, OFA and **EBA**, represent faces and bodies independently. The face and body areas in the fusiform gyrus (FFA, **FBA**)

Glossary

Biological motion: motion generated by biological agents, such as humans and animals. It involves different neural and cognitive mechanisms than motion generated by nonbiological stimuli.

Changeable facial aspects: faces convey very rich information that varies from moment to moment. This includes facial expressions, mouth and eye movements. Current neural models posit that the STS mediates the processing of changeable facial aspects consistent with their important role in social communication [54].

Dynamic identity signature: identity cues that reside in idiosyncratic elements of face and body movements. These are nuanced, person-specific expressive gestures that punctuate social and pragmatic face and body movements. For these dynamic identity signatures to facilitate person recognition, the visual system must represent and remember, not only the categorical content of biological motions (e.g., smiling, walking, getting out of a car), but also person-specific variations of these; that is, the way he smiles, walks, or gets out of a car.

Extrastriate body area (EBA): a body-selective brain area in the lateral occipital cortex [61]. The EBA is defined functionally within each individual as the cluster of voxels in the extrastriate cortex that shows significantly higher response to body than to object stimuli.

Form-from-motion: information from faces and bodies can benefit person recognition by enhancing the quality of shape representations of a face or person. These cues are rooted in basic perceptual abilities to see structure from motion, where shape is revealed more clearly through motion.

Fusiform body area (FBA): a body-selective brain area in the middle of the fusiform gyrus. The FBA is defined functionally within each individual as the cluster of voxels in the fusiform gyrus that shows significantly higher response to body than to object stimuli [62].

Fusiform face area (FFA): a face-selective brain area in the middle of the fusiform gyrus [63]. The FFA is defined functionally within each individual as the cluster of voxels in the fusiform gyrus that shows