Recognizing moving faces:
A psychological and neural synthesis

Alice J. O'Toole*, Dana A. Roark, Hervé Abdi
The University of Texas at Dallas

March 13, 2002

Key words: facial motion, face recognition, dynamic identity signatures, neural processing of faces, face perception, facial expression

*Address for all authors:
A. J. O'Toole
School of Human Development, GR4.1
University of Texas at Dallas
Richardson TX 75083
otoole@utdallas.edu

Summary

Information for identifying a human face can be found both in the invariant structure of features and in idiosyncratic movements and gestures. When both kinds of information are available, psychological evidence indicates that: 1.) dynamic information contributes more to recognition under non-optimal viewing conditions, e.g., poor illumination, low image resolution, recognition from a distance; 2.) dynamic information contributes more as a viewer's experience with the face increases; and 3.) a structure-from-motion analysis can make a perceptually-based contribution to face recognition. A recently proposed distributed neural system for face perception, with minor modifications, can accommodate the psychological findings with moving faces.

A furrowed brow, a tilt of the head, a roll of the eyes, … faces "speak" by movements, as well as words. Indeed, facial motion is vital to human communication [see Box 1], but does it help or hinder the recognition process? Can visual *memory* use dynamic information for face recognition? We first review the literature on recognition of moving faces and organize the common themes that emerge into explicit hypotheses. Next, we link current knowledge about the neural basis of face processing to the behavioral findings. Finally, we propose a framework for mapping studies of memory for moving faces onto plausible neural processing systems.

Hypotheses

In the current literature, there are two hypotheses about the role of facial motion in recognition. The *supplemental information hypothesis* posits that we represent characteristic facial motions or gestures of individual faces, in addition to the invariant structure of the face[1-7]. We refer to these idiosyncratic movements as *dynamic facial signatures*[5] [see Box 2 for a description]. When both static and dynamic identity information are available, people are likely to rely primarily on static information for recognition. This is because dynamic facial signatures probably provide less reliable identification information than static facial structure [see Box 2].

The *representation enhancement hypothesis* posits that facial motion contributes to recognition by facilitating the perception of the three-dimensional structure of a face [8,9]. This hypothesis assumes that motion adds to the quality of the structural information accessible from a human face, and that this benefit transcends the additional views of the face provided from the motion**.** The rationale behind this hypothesis draws implicitly on structure-from-motion phenomena, which have been studied extensively in psychology and in computer vision[10,11], [See Box 3].

The supplemental information and representation enhancement hypotheses are not mutually exclusive, but rather, may be complementary. The former suggests a direct encoding of idiosyncratic *facial movements*. The latter suggests that motion contributes to a perceptual process that enhances the encoding of the *invariant structure* of the face.

Psychological Studies of Motion and Face Recognition

With a brief glance across a dimly lit room, we can recognize the face of a friend. For unfamiliar faces, however, even relatively minor changes in viewpoint[12] and illumination[13] impair performance[14,15]. The importance of familiarity in supporting robust recognition has been shown also for moving faces and people. For example, Burton et al. found that subjects can identify familiar, but not unfamiliar, people from poor-quality surveillance video[16]. Further, even when high-quality video is used, recognition for unfamiliar faces falters when people are asked to pick out the target face from displays that contain similar-looking distracter faces[17,18]. Consequently, the studies we review are best divided into those employing familiar faces, which address the supplemental information hypothesis, and those using unfamiliar faces, which address the representation enhancement hypothesis.

Familiar Faces

Research with familiar/famous faces consistently supports the supplemental information hypothesis. Motion facilitates familiar/famous face recognition under a variety of non-optimal recognition conditions. Participants in these studies are asked typically to name famous faces from spatially degraded motion displays. The methods used to distort/degrade facial images include the use of photographic negatives, inversion, image threshholding, blurring, and pixelating[4,5,7]. Moreover, the quality or "naturalness" of the motion appears to contribute to the *amount* of benefit derived from motion. Slowing down or disrupting the rhythm of faces impairs recognition by comparison to evenly-paced moving faces[5,6]. Recognition can even be "primed" or facilitated by pretest exposure to moving faces[19].

Unfamiliar Faces

To date, it is unclear whether motion benefits unfamiliar face recognition. In most studies, subjects view previously unfamiliar faces or people from either dynamic or static displays during learning and are tested subsequently with dynamic or static displays. We consider both eyewitness studies, using whole bodies and events as stimuli, and face recognition studies.

For the eyewitness experiments, two early studies found conflicting results. Schiff and colleagues found a motion advantage when participants viewed a staged robbery and were tested with "dynamic mug shots" [20]. These mug shots showed faces rotating through 180-degrees from left to right profile. However, no benefit was found for learning from the dynamic presentation over viewing the robbery video from two freeze frames. Using a similar eyewitness approach, Shepherd and his colleagues found that videotaped presentations conferred no benefit over single photographs at either learning or test, although "live" presentations of suspects elicited the best recognition[21].

There are also conflicting results with unfamiliar face recognition studies. Pike and colleagues found beneficial effects of facial motion[8]. Learning from dynamic displays of heads rigidly rotating over 360 degrees fared better than learning from static images taken from multiple viewpoints. Christie and Bruce, however, did not replicate these results. They found no benefit for the moving stimuli, either at learning or at test[9]. A slightly different task that involved participants matching a previously viewed face to one presented in an array of several faces, found no recognition benefit for dynamic faces[17]. However, Thornton and Kourtzi showed that video clips of unfamiliar faces presented as "primes", improved subjects' subsequent accuracy for matching the primes to target faces[22]. Additionally, Bruce et al. found that social interaction between participants (i.e., conversing about the faces during learning) improved face identification in a match-to-sample paradigm[23].

In summary, although facial motion information can benefit *familiar* face recognition, its effects for relatively *unfamiliar* faces are less certain. Either motion is not helpful with unfamiliar faces, or its benefits are offset by other factors. The social significance of facial motions [see Box 1], which may distract the viewer from the task of encoding the identity of a face, is one such factor. We will discuss this shortly.

Neural Basis of Face Perception

The neural basis of face processing has been studied for decades and has yielded findings concerning the organization of high-level visual brain areas responsive to various aspects of faces. Many studies have noted the various cortical regions involved in processing faces[24,25]. Recently, based on the broad activation of cortical areas in response to faces, Haxby and his colleagues proposed a "distributed" neural system for face processing that distinguishes between the representation of the invariant versus changeable aspects of faces[26]. We focus on this particular distributed model because it may give insight into the psychological data on memory for moving faces. Haxby et al. posit two core areas along with "extender regions" for specific tasks. For the core system, the lateral fusiform gyrus (also called the fusiform face area, FFA), active in many neuroimaging studies of face perception, represents invariant facial information useful for identifying faces. (The analogous general system in primates is the inferotemporal, IT, cortex.) The posterior superior temporal sulcus (pSTS) processes the changeable aspects or movements of faces. Neurophysiological studies in non-human primates and neuroimaging studies in humans indicate that this area is important for detecting gaze information, head orientation, and expression[26-29]. Biological motion of the whole body, hand, eyes, and mouth also activates the superior temporal sulcus (STS)[28].

In addition to the core system, the STS system is extended to include brain areas involved in gaze and orientation detection, speech perception from mouth movements, and the perception of emotion. The IT system is extended to brain areas involved in the retrieval of personal identity, name, and biographical information[26].

The functional division of these high level brain areas for processing dynamic versus invariant information about faces is consistent with the channeling of visual information beginning at the retina, into the high resolution, color sensitive, *parvocellular* stream, and the lower resolution, motion sensitive, *magnocellular* stream[30]. These inputs map onto the "what" (ventral) and "where" (dorsal) visual streams, which guide visual object recognition and spatial orientation, respectively[31]. Thus, dynamic information from faces may be carried primarily in the dorsal stream, whereas the static features of faces may be processed best in the ventral stream.

Memory for Moving Faces: The Interplay of Two Systems

The physical and functional divisions associated with the changeable and invariant brain systems for face processing return us to the psychological studies. Given that both the changeable and static aspects of faces are useful for identification, and that the former is processed in the STS and the latter in IT, how do these systems interact when we recognize a moving face? Additionally, if facial motions are also critical carriers of social communication signals, do these signals catch our attention and help us focus on identity? Or, does the added burden of processing social communication information from moving faces distract us from attending to identity?

A fascinating example of the dissociable nature of social and identity processing was provided recently by Simons and Levin[32]. They observed close-range interactions between a naïve subject and an experimenter who posed as a construction worker asking for directions. These interactions were interrupted briefly by construction workers carrying a door between the subject and experimenter. During this brief separation, the experimenter changed places with a different experimenter. An astounding 60 percent of the naïve subjects failed to notice the "person" change, and continued to give directions as if nothing had happened**.** Under some circumstances, therefore, it would seem that the processing of social and identity information can proceed independently [See Box 4].

We think that the separate streams hypothesis for changeable social and invariant identity information[26] could serve as a useful framework for understanding human memory for moving faces. This model can accommodate the human memory data with two minor modifications [see Figure 1]. In addition to the primary face recognition system that processes static information along the ventral stream to IT, a secondary system that processes dynamic identity information along the dorsal stream to STS seems likely. In our hypothesized model, dynamic signatures, which are embedded in expressions, facial speech, and orienting head/face movements, are processed in the dorsal stream, transiting through the general visual areas that support motion processing (e.g., MT) to the STS. Aspects of dorsal-stream processing, like its ability to operate in poor illumination and with low resolution stimuli, make it ideal as a possible secondary route to recognition for stimuli that are not "ventral system quality." To use this secondary face recognition system, however, we must know a face well enough to have learned its idiosyncratic movements. Despite the relatively modest amount of data on dynamic facial signatures, we wonder if this system supports the robust illumination and view-invariant face and person recognition we exhibit for familiar people. So, the first modification of the distributed model amends the role of STS stream to include identification, when identification information is inherent in the motion of the face.

The second modification is more speculative and concerns structure-from-motion analyses for face recognition. Facial motion from both familiar and unfamiliar faces could serve as input to this analysis. Information for this analysis might proceed through the dorsal stream to MT and ultimately back to IT, but as *static form* information [i.e., "motionless form", see Fig. 1]. Indeed, neurons in primate IT, sensitive to particular forms, respond invariantly to form even when it is specified by pure motion-induced contrasts[33]. Lesion studies also indicate that form discrimination mechanisms in IT can make use of input from the motion processing system[34]. Both the neurophysiological[33] and lesion studies[34] suggest known connections from MT to IT via V4 [35,36] as a plausible basis of their findings. At present, however, psychological demonstrations of the usefulness of this route for face recognition have been indirect [see Box 3].

Conclusions

When both static and dynamic information about facial identity are available, the psychological evidence points to the following principles in balancing the contributions of the two kinds of input:

- **Dynamic information contributes more in poor viewing conditions.** Because facial structure is a more reliable cue to recognition than the dynamic identity signature, motion information is most beneficial when viewing conditions are not optimal for extracting the facial structure.
- **Face familiarity mediates the role of dynamic information in recognition**. Because characteristic motions and gestures occur only intermittently, they are learned more slowly than static facial features. The relative importance of motion information to recognition, will increase, therefore, with a viewer's experience with the face.
- **Facial motion can contribute to a perceptual structure-from-motion analysis.** Facial motion can bootstrap the encoding of the invariant facial structure, by providing information that is unavailable from the pictorial cues alone. The quality of the viewing conditions may mediate the contribution of this structure-from motion analysis to recognition. Familiarity with the face, however, is not a prerequisite.

These principles, combined with the distributed model, provide a viable framework for synthesizing the disparate psychological results about the role of motion in memory for faces.

References

1 Bruce, V. and Valentine, T. (1988) When a nod's as good as a wink: the role of dynamic information in facial recognition. *Pract. Aspects Mem. Curr. Res. Ideas* 1, 169-174

2 Hill, H. and Johnston, A. (2001) Categorizing sex and identity from the biological motion of faces. *Curr. Biol*. 11, 880-885

3 Knappmeyer, B. *et al.* (2001) Facial motion can determine identity. *J. Vis.* 3, 337

4 Knight, B. and Johnston, A. (1997) The role of movement in face recognition. *Vis. Cognit*. 4, 265-273

5 Lander, K. *et al,* (1999) The role of movement in the recognition of famous faces. *Mem. & Cognit*. 27, 974-985

6 Lander, K. and Bruce, V. (2000) Recognizing famous faces: exploring the benefits of facial motion. *Ecol. Psychol.* 12, 259-272

7 Lander, K. *et al*. (2001) Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Appl. Cog. Psychol.* 15, 101-116

8 Pike, G.E. *et al.,* (1997) Recognizing moving faces: the relative contribution of motion and perspective view information. *Vis. Cognit*. 4, 409-437

9 Christie, F. and Bruce, V. (1998) The role of dynamic information in the recognition of unfamiliar faces. *Mem. Cognit*. 26, 780-790

10 Johansson, G. (1973) Visual perception of biological motion and a model for its analysis. *Perc. Psych*. 14, 201-211

11 Ullman, S. (1979) *The Interpretation of Visual Motion.* Cambride, MA : MIT Press

12 O'Toole, A. J., *et al,* (1998) Stimulus-specific effects in face recognition over changes in viewpoint. *Vis. Res*. 38, 2351-2363.

13 Hill, H. and Bruce V. (1996) Effects of lighting on the perception of facial surface. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 986-1004

14 Hancock, P.J.B. et al. (2000) Recognition of unfamiliar faces. *Trends Cognit. Sci.* 4, 330-337

15 Burton, A.M. *et al.,* (1999) From pixels to people: a model of familiar face recognition. *Cognit. Sci*. 23, 1-31

16 Burton, M.A. *et al*., (1999) Face recognition in poor-quality video. *Psychol. Sci.* 10, 243-248

17 Bruce, V. *et al.,* (1999) Verification of face identities from images captured on video. *J. Exp. Psychol. Appl.* 5, 339-360

18 Henderson, Z. *et al.,* (2001) Matching the faces of robbers captured on video. *Appl. Cog. Psychol*. 15, 445-464

19 Lander K. and & Bruce V (2001) Repetition priming of moving faces. Talk presented at the Euro. Conf. Of Cog. Psych., ESCOP) X, Edinburgh Scot., Sept. 2001

20 Schiff, W. *et al.,* (1986) Recognizing people seen in events via dynamic "mug shots." *Amer. J. Psychol*. 99, 219-231

21 Shepherd, J.W. *et al.,* (1982) *Identification Evidence: A Psychological Evaluation*, University of Aberdeen Press

22 Thornton, I.M. and Kourtzi, Z. A matching advantage for dynamic human faces. *Perception* (in press)

23 Bruce, V. et al. (2001) Matching identities of familiar and unfamiliar faces caught on CCTV images. *J. Exp. Psy.: Appl*. 7, 207-218

24 Damasio, A.R. et al. (1982) Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology* 32, 331-341

25 Sergent, J. et al. (1992) Functional neuroanatomy of face and object processing: A Positron emission tomography study. *Brain* 115, 15-36

26 Haxby, J.V. et al. (2000) The distributed human neural system for face perception. *Trends Cognit. Sci.* 4, 223-233

27 Haxby, J.V. et al (2002) Human neural systems for recognition and social communication. *Biol. Psychiatry* 51, 59-67.
28 Allison, T. et al. (2001) Social perception from visual cues: role of the STS region. *Trends Cognit. Sci.*4, 267-278
29 Narumoto, J. et al. (2001) Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Cog. Brain Res*. 12, 225-231
30 Merigan, W. H. (1991) P and M pathway specialization in the macaque. In *From Pigments to Perception*. (A. Valberg and B.B. Lee, eds), pp. 117-125, Plenum Press.
31 Ungerlieder, L.G. and Mishkin, M. (1982) Two cortical visual systems. *In Analysis of Visual Behavior* (Ingle, D.J. et al., eds), pp. 549-586, MIT Press.
32 Simons, D.J. and Levin, D. T. (1998) Failure to detect changes to people during a real-world interaction. *Psychonomic Bull. Rev* 5 644-649
33 Sary G., et al. (1993) Cue invariant shape selectivity of macaque inferior temporal neurons. *Science* 260, 995-997
34 Britten K.H., et al. (1992) Effects of inferotemporal lesions on form-from-motion discrimination. *Exp. Brain Res*. 88, 292-302
35 Maunsell, J.H.R. and Van Essen, D.C. (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J Neurophysiol*. 3, 2563-2586
36 Ungerleider, L.G. and Desimone, R. (1986) Cortical connections of visual area MT in the macaque. J. Comp. Neurol. 248, 190-222

Box 1. Facial motion: "What is it?" and "What is it for?"

Most *rigid head motions* can be interpreted as social interaction signals. To begin or end a conversation we turn our heads to look at, or away from, someone. We redirect the attention of others with a head-turn and we nod to indicate agreement. Rigid head movements provide the observer with a moving stimulus and with more views of the head than would be encountered from a static observer and subject. Psychological studies have distinguished, therefore, between recognition effects due to the number of views seen and those due to the motion of the head [e.g., Refs a,b].

Fig. I. Rigid head movements such as rotations provide more views of the head than are available from a static head.

Non-rigid head movements can be grouped into speech production movements, facial expression movements, and eye gaze changes. The visual cues provided by the face during speech function to boost the intelligibility of speech [e.g., Refs c,d,e,f]. Facial expressions can convey a person's mood [Fig. II]. Changes in the direction of eye gaze provide information about the object of attention [Ref g].

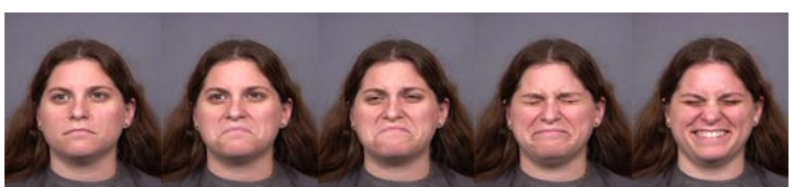

Fig. II. Non-rigid facial movements can convey the unfolding of a facial expression, as in this figure where the expression goes from a neutral expression to disgust.

In natural situations, such as during a conversation, rigid and non-rigid movements are combined to communicate more complex information [Fig. III].

Fig. III. Combinations of rigid and non-rigid facial motions are common. The woman in this sequence is simultaneously speaking, smiling, shifting her gaze, and tilting her head back and forth rigidly.

References

a Christie, F. and Bruce, V. (1998) The role of dynamic information in the recognition of unfamiliar faces. *Mem. & Cog.* 26, 780-790

b Pike, G.E. et al. (1997) Recognizing moving faces: the relative contribution of motion and perspective view information. *Vis. Cognit.* 4, 409-437

c Rosenblum, L.D. et al. (1996) Point-light facial displays enhance comprehension of speech in noise. *J Speech Hearing Res.* 39, 1159-1170

d McGurk, H. and MacDonald, T. (1976) Hearing lips and seeing voices. *Nature* 264, 746-748

e Calvert, G.A. *et al.,* (1998) Crossmodal identification. *Trends Cog. Sci*.,2, 247-253

f Campbell, R. *et al,* eds. *(1998)* Hearing by Eye II: Advances in the Psychology of Speechreading Auditory Vis. Speech*, Psychology Press*

g Langton, S.R.H. et al. (2000) Do the eyes have it? Cues to the direction of social attention*. Trends Cog. Sci.* 4, 50-59

Box 2 Facial motion and Identity: Dynamic Signatures

Fig. I. Projections of human facial movements onto synthetic heads used as stimuli by Hill and Johnston [Reproduced with permission from Ref. f.]

Dynamic identity signatures [Refs a-d] refer to characteristic or "identity-specific" movements of the face, head, or body. Are these movements useful for identifying a person? Logically, identity-specific facial motion is likely to be less reliable than information about the structure of a face. It is possible with a single static encounter to encode a person's distinctive nose. After viewing a distinctive facial gesture, however, we cannot know if the gesture is idiosyncratic. Even with enough familiarity to know a face's dynamic identity signature, we cannot rely on these gestures being repeated reliably at future encounters.

Consequently, dynamic facial signatures may be learned more slowly than facial features, and may become *relatively* more reliable for identification when we know the person better. Indeed, psychological studies suggest that we rely on motion information to identify familiar faces only when the featural cues are not easily accessible [Refs a-d].

The role of characteristic motions for recognition of unfamiliar faces has been investigated recently using animated synthetic 3D head models [Ref. e]. These studies focus on the "learnability" of a dynamic signature when it is the most, or only, reliable cue to identity [Refs f,g]. For example, Hill and Johnston projected facial animations generated by human actors onto a computer-generated average head [Ref. f]. Subjects learned to discriminate individuals based solely on the facial motion information. Knappmeyer and her colleagues trained participants to discriminate two synthetic faces animated with different characteristic facial motions [Ref. g]. When later viewing morphs between the two head models, identity judgments about the intermediate morphed heads were biased by the animated motion information associated with the faces originally. Both studies show that dynamic information from face movements can support recognition.

References

a Knight, B. and Johnston, A. (1997) The role of movement in face recognition. *Vis. Cognit.* 4, 265-273

b Lander, K. et al. (1999) The role of movement in the recognition of famous faces. *Mem. & Cognit.* 27, 974-985

c Lander, K. and Bruce, V. (2000) Recognizing famous faces: exploring the benefits of facial motion. *Ecol. Psy.* 12, 259-272

d Lander, K. et al. (2001) Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Appl Cog. Psy.* 15, 101-116

e Blanz, V. & Vetter, T. (1999) A morphable model for the synthesis of 3D faces. SIGGRAPH 99 Proceedings, Los Angelos, CA.

f Hill, H. and Johnston, A. (2001) Categorizing sex and identity from the biological motion of faces. *Curr. Biol.* 11, 880-885

g Knappmeyer, B. *et al.* (2001) Facial motion can determine identity. *J. Vis. 3*, 337.

Box 3. Structure-from-motion in the perception of faces and bodies

The structure of human faces can be perceived from “point light displays” [Ref. a]. These displays isolate motion from the pictorial cues that normally accompany it. The age of a speaker [Ref. b], as well as information about facial expression [Refs c-d], can be extracted from these displays. Seeing point light faces can even facilitate the perception of speech [Ref. e]. The quality of *identity* information available from point light displays of faces, however, is less certain. Although Bruce and Valentine found that subjects could recognize friends’ faces from moving point light displays, accuracy was quite poor [Ref. f].

Interpreting point light studies of faces in the context of the supplemental information and representation enhancement hypotheses is limited by the fact that dynamic identity signature information [see Box 2] and structure-from-motion information are intermixed in these displays. For example, studies on the perception of gender from point light walkers have shown that gender judgments are supported both by the animation of the walker, which reveals static body structure (e.g., center of gravity and shoulder-to-hip ratio), and by characteristically dynamic male and female walking styles (e.g. hip swing), [Ref g]. Also, Hill and Pollick showed that the temporal properties of characteristic body movements directly affect identification. When asked to identify individuals based on signature arm movements, participants’ recognition accuracy increased when they saw temporally exaggerated point-light sequences as compared to the originally learned, unexaggerated sequences [Ref. h].

For identification of faces, point light displays support some limited ability to process dynamic identity signatures and possibly also the structural information underlying the motion. At present, however, there is no direct evidence for sorting through the structure versus dynamic signature cues in the perception of facial identity. Recent studies using animated computer-based head models, where the identity-specific form information in a face is ambiguous or modifiable [Refs. i, j; see Box 2], isolate the dynamic signature information, but cannot provide a pure measure of the structure-from-motion information.

References

a Johansson, G. (1973) Visual perception of biological motion and a model for its analysis. *Perc. & Psychophysics* 14, 201-211

b Berry, D.S. (1990) What can a moving face tell us? *J. Pers. Soc. Psy.* 58, 1004-1014

c Bassilli, J.N. (1978) Facial motion in the perception of faces and emotional expression. *J. Exp. Psych.* . 4, 373-379.

d Bassilli, J.N. (1979) Emotion recognition: the role of facial movement and relative importance of upper and lower areas of the face. *J. Pers. Soc. Psych*. 37, 2049-2058.

e Rosenblum, L. D. and Saldana, H.M. An audiovisual test of kinematic primitives for visual speech perception. *J. Exp. Psych. Hum. Perc. Perf*. 22, 318-331.

f Bruce, V. and Valentine, T. (1988) When a nod's as good as a wink: the role of dynamic information in facial recognition. *Pract. Aspects Mem. Curr. Res. Ideas* 1, 169-174

g Stevenage, S. et al. (1998) Visual analysis of gait as a cue to identity. *Appl. Cog. Psychol*. 13, 513-526

h Hill, H. and Pollick, F.E. (2000) Exaggerating temporal differences enhances recognition of individuals from point light displays. Psychol. Sci. 3, 223-228.

i Hill, H. and Johnston, A. (2001) Categorizing sex and identity from the biological motion of faces. *Curr. Biol. 11*, 880-885

j Knappmeyer, B., *et al.* (2001). Facial motion can determine identity. *J. Vis. 3, 337*

Box 4. Expression and Identity: A model for motion processing?

The processing of facial expression and identity has long been posited to be functionally distinct [Ref. a]. Neuropsychological double dissociations of expression and identity are consistent with this claim. Indeed, some prosopagnosics, who are unable to recognize faces, nonetheless retain the ability to perceive facial expressions accurately [Refs. b,c], and some patients with an impaired ability to process facial expression retain the ability to identity faces [e.g., Ref. d].

Psychological studies using the Garner speeded classification task support an "asymmetric" dissociation for expression and identity, with the expression analysis accessing some identity information, but the identity analysis proceeding independently [Refs. e,f]. The Garner task was developed to test the (in)dependence of stimulus dimensions [Ref. g]. Two stimulus dimensions are considered independent if people can attend selectively to either dimension, while ignoring irrelevant variations in the other. Although early evidence indicated mutual independence for expression and identity [Ref. h], recent work suggests an "asymmetric dependence" [Refs. e,f]. Specifically, identity judgments are not influenced by irrelevant changes in expression, but expression judgments are influenced by irrelevant variations in identity [Ref. e]. A similar result occurs for processing identity and facial speech [Ref. f].

Computational models of facial expression analysis support the underlying premise of processing independence by illustrating that facial variations relevant to identity versus those relevant for expression can be separated physically [Ref. i]. A principal component analysis of faces varying in identity and expression revealed that expression and identity were largely coded by different sets of principal components. Thus, independence could emerge naturally from the statistical structure of the faces.

References

a Bruce, V. and Young, A. (1986) Understanding face recognition. *Brit. J. Psych*. 77, 305-27

b Kurucz, J. Feldmar, G. (1979) Prosopo-affective agnosia as a symptom of cerebral organic brain disease. *J. Amer. Geriatrics Soc.* 27, 91--95.

c Kurucz, J., Feldmar, G. & Werner, W.,(1979). Prosopo-affective agnosia associated with chronic organic brain syndrome. *J. Amer. Geriatrics Soc.* 27, 225--230.

d Bruyer, R, et al. (1983) A case of prosopagnosia with some preserved covert remembrance of similar faces*. Brain and Cognit.* 2, 257--284.

e Schweinberger, S.R. et al. (1999) Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perc. Psych.* 61, 1102-1115 f Schweinberger, S.R. and Soukup, G.R. (1998) Asymmetric relationships among perceptions of facial identity, emotion and facial speech. *J. Exp. Psych. Hum. Perc. Perf*. 24, 1748-1765

g Garner, W. R. (1974) *The processing of information and structure*. Lawrence Erlbaum.

h Etcoff, N.L. (1984) Selective attention to facial identity and facial emotion. *Neuropsych*. 22, 281-195

i Calder, A.J., et al. (2001). A principal component analysis of facial expressions. *Vis. Res.* 41, 1179--1208.

Outstanding Questions

- How do different *types* of facial movement affect face recognition?
- Does the social relevance of facial movements affect recognition?
- How much, and what kind, of experience with a face is necessary to be able to recognize someone based on their characteristic pattern of facial movement?
- Could the recognition of dynamic signatures from faces occur in the FFA? Is the face recognition system in FFA motion blind?

- Figure 1

Title: Mapping recognition of moving faces onto neural systems: A model

Caption:

We propose Haxby et al.'s distributed neural system for face perception[26] as a framework for understanding psychological findings regarding the effects of facial motion on memory for faces. The ventrally-based stream (blue) processes the static structure of a face, and the dorsally-based stream (red) processes facial motion. Facial motion contains two different types of information: social communication signals (gaze, expression and facial speech) and person-specific dynamic facial signatures. Following Haxby et al., the social communication information is forwarded to STS and then to the extender systems responsible for specific social tasks. We suggest that dynamic facial signatures embedded in expression, facial speech and gaze are processed in STS as well and can provide a secondary route to face recognition for familiar faces (supplemental information hypothesis). We may rely on this secondary system for recognition when the viewing conditions are non-optimal (low illumination, poor resolution, recognition from a distance). We speculate also that structure-from-motion may benefit face recognition via communication between the dorsal and ventral streams. Plausibly, information from MT might contribute to the structural representation of a face in IT/FFA (see text). This added input could benefit recognition for either familiar or unfamiliar faces (representation enhancement hypothesis).

Acknowledgements

Thanks are due to Jim Bartlett, Josh Harms, Dawn Hurst, Jim Jerger, George Lannen, Matt Pappas, Sarah Snow, and Anneliese West, for useful discussions of these issues, helpful comments on the manuscript, and assistance with figure preparation. Thanks are due also to David Leopold, Jon Pauls, and David Scheinberg and to Stephen Lomber for guidance on the neuroscience aspects of the paper. This work was supported under the Human Identification Project (DOD/DARPA) by a grant to A.O'T. and H.A.