Something in the Way We Move: Motion Dynamics, not Perceived Sex, Influence Head Movements in Conversation

Steven M. Boker
University of Virginia

Jeffrey F. Cohn
Barry–John Theobald
Iain Matthews
Michael Mangini
Jeffrey R. Spies
Zara Ambadar
Timothy R. Brick

Draft March 27, 2009
Please do not cite or quote

Abstract

During conversation, women tend to nod their heads more frequently and more vigorously than men. An individual speaking with a woman tends to nod his or her head more than when speaking with a man. Is this due to social expectation or due to coupled motion dynamics between the speakers? We present a novel methodology that allows us to randomly assign apparent identity during free conversation in a videoconference, thereby dissociating apparent sex from motion dynamics. The method uses motion–tracked synthesized avatars that are accepted by naive participants as being live video. We find that 1) motion dynamics affect head movements but that apparent sex does not; 2) judgments of sex are driven almost entirely by appearance; and 3) ratings of masculinity and femininity rely on a combination of both appearance and dynamics. Together, these findings are consistent with the hypothesis of separate perceptual streams for appearance and biological motion. In addition, our results are consistent with a view that head movements in conversation form a low level perception and action system that can operate independently from top–down social expectations.

Introduction

When humans converse, we adapt multimodally to one another. Semantic content of conversation is accompanied by vocal prosody, non–word vocalizations, head movement, gestures, postural adjustments, eye movements, smiles, eyebrow movements, and other facial muscle changes. Coordination between speakers' and listeners' head movements, facial expressions, and vocal prosody has been widely reported (Bernieri, Davis, Rosenthal, & Knee, 1994; Cappella, 1981; Condon, 1976; Lafrance, 1985). Conversational coordination can be defined as when an action generated by one individual is predictive of a symmetric action by another (Rotondo & Boker, 2002; Griffin & Gonzalez, 2003). This coordination is a form of spatiotemporal symmetry between individuals (Boker & Rotondo, 2002) that has behaviorally useful outcomes (Chartrand, Maddux, & Lakin, 2005).

Head movements, facial expressions, and vocal prosody influence our perceptions of other people, including features such as identity (Lander, Christie, & Bruce, 1999; Munhall & Buchan, 2004), rapport (Grahe & Bernieri, 2006; Bernieri et al., 1994), attractiveness (Morrison, Gralewski, Campbell, & Penton-Voak, 2007), gender (Morrison et al., 2007; Hill & Johnston, 2001; Berry, 1991), personality (Levesque & Kenny, 1993), and affect (Hill, Troje, & Johnston, 2003). Point light displays have been used to show that affective information can be transmitted via motion cues from gestures (Atkinson, Tunstall, & Dittrich, 2007; Clarke, Bradshaw, Field, Hampson, & Rose, 2005) and facial expressions (Pollick, Hill, Calder, & Paterson, 2003). In dyadic conversation, each conversant's perception of the other person produces an ever–evolving behavioral context that in turn influences her/his ensuing actions, creating a nonstationary dynamical system with feedback as conversants form patterns of movements, expressions, and vocal inflections; sometimes with high symmetry between the conversants and sometimes with little or no similarity (Ashenfelter, Boker, Waddell, & Vitanov, in press; Boker & Rotondo, 2002).

The context of a conversation can influence its course: The person you think you are speaking with can influence what you say and how you say it. The appearance of an interlocutor is composed of his/her facial and body structure as well as how he/she moves and speaks. Judgments of gender for perception of static facial images rely on information that is to a large degree concentrated around the eyes and mouth (Mangini & Biederman, 2004; Schyns, Bonnar, & Gosselin, 2002) as well as cues from hair (Macreae & Martin, 2007). Judgments of gender can also be made entirely from facial dynamics — expressions and rigid head motion (Hill & Johnston, 2001). Berry (1991) presented silent point light faces engaging in interaction and reciting passages. Adults could recognize gender from the faces at greater than chance levels in both, but children could only recognize gender in the

This work was supported by NSF Human and Social Dynamics Grant BCS–0527485 and EPSRC Grant EP/D049075/1. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also gratefully acknowledge the help of Kathy Ashenfelter, Tamara Buretz, Eric Covey, Pascal Deboeck, Katie Jackson, Jen Koltiska, Sean McGowan, Sagar Navare, Stacey Tiberio, Michael Villano, and Chris Wagner. We also acknowledge the substantial help of three thoughtful and detailed reviewers. Correspondence may be addressed to Steven M. Boker, Department of Psychology, The University of Virginia, PO Box 400400, Charlottesville, VA 22903, USA; email sent to boker@virginia.edu; or browsers pointed to http://people.virginia.edu/˜smb3u.

interacting faces, suggesting that gender–based motion cues during conversation are stronger than during acted sequences. One gender difference in dynamics during conversation is that women tend to use more nonverbal backchannel cues (Duncan & Fiske, 1977; Roger & Nesshoever, 1987) including nodding their heads more often (Helweg-Larsen, Cunningham, Carrico, & Pergram, 2004) and more vigorously (Ashenfelter et al., in press) than do men. But who one is speaking with also matters: A person talking to a woman tends use more backchannel cues (Dixon & Foster, 1998) and to nod more vigorously than he/she nods when talking to a man (Ashenfelter et al., in press).

Cognitive models with separate pathways for perception of structural appearance and biological motion have been proposed (Giese & Poggio, 2003; Haxby, Hoffman, & Gobbini, 2000). Evidence for this view comes from neurological (Humphreys, Donnelly, & Riddoch, 1993; Steede, Tree, & Hole, 2007b, 2007a) and judgment (Knappmeyer, Thornton, & Bülthoff, 2003; Hill & Johnston, 2001) studies. The phenomenon of increased head nodding when speaking with a woman might be due to her appearance, i.e., a form of social expectation, or it might be due to dynamic coupling to driven by perception of biological motion.

It is difficult to separately manipulate the static and dynamic influences on conversational context. An individual conversant has a characteristic and unified appearance, head motions, facial expressions, and vocal inflection. For this reason, most studies of person perception and social expectation are naturalistic or manipulations in which behavior is artificially scripted and acted. But scripted and natural conversation have differences in dynamic cues (Berry, 1991). We present a novel methodology that allows manipulation of appearance using a real–time resynthesized near–photorealistic avatar such that conversants can carry on conversations without knowing which sex their interlocutors perceive them to be.

We present the results of three experiments. The first experiment tests naive participants' perceptions of believability and naturalness of the avatar faces in free–form conversation. The second experiment changes the apparent sex of one participant in a dyadic conversation and tracks head movements of both participants to test whether gender differences in head nodding behavior are related to apparent sex or by dynamics of head movements, facial expressions, and vocal prosody. The third experiment tests whether the apparent sex of the avatar is convincing by re–rendering 10s clips of conversation from the second experiment as both male and female avatars and then displaying the rendered clips to naive raters.

Materials and Methods

We are presenting a novel methodology for manipulating perceptions during dyadic perception and action experiments as well as results from the application of that methodology. We begin by describing the methods used to create the resynthesized avatars and apply the expressions of one person onto the appearance of another — methods and apparatus that is common to the three experiments.

Resynthesized Avatars

The resynthesised avatars used in our work are based on Active Appearance Models (AAMs) (Cootes, Edwards, & Taylor, 2001; Cootes, Wheeler, Walker, & Taylor, 2002;

Matthews & Baker, 2004). AAMs provide a compact statistical description of the variation in the shape and the appearance of a face. The shape of an AAM is defined by the vertex locations,

$$\mathbf{s} = \{x_1, y_1, x_2, y_2, \ldots, x_n, y_n\}^T,$$

of a two-dimensional (2D) triangulated mesh that delineates the facial features (eyes, mouth, etc.). The topology of the mesh (number and interconnection of the vertices) is fixed, but the vertex locations undergo both rigid (head pose) and non-rigid (facial expression) variation under the control of the model parameters. The appearance of the AAM is an image of the face, which itself varies under the control of the parameters.

To construct an AAM a set of training images is required. These are images chosen to represent the characteristic variation of interest, for example prototypical facial expressions. The triangulated mesh is overlaid onto the images and the vertex locations are manually adjusted so that they align with the facial features in the images. The set of N training shapes is represented in matrix form as $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N]$, and principal components analysis (PCA) applied to give a compact model of the form:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{m} \mathbf{s}_i p_i, \tag{1}$$

where $\mathbf{s}_0$ is the mean shape and the vectors $\mathbf{s}_i$ are the eigenvectors of the covariance matrix corresponding to the m largest eigenvalues (see Figure 1–a). These eigenvectors are the basis vectors that span the shape-space and they describe changes in the shape relative to the mean shape. The coefficients p_i are the shape parameters, which define the contribution of each basis in the reconstruction of $\mathbf{s}$. An alternative interpretation is that the shape parameters are the coordinates of $\mathbf{s}$ in shape-space, thus each coefficient is a measure of the distance from $\mathbf{s}_0$ to $\mathbf{s}$ along the corresponding basis vector.

The appearance of the AAM is a description of the pixel intensity variation estimated from a shape-free representation of the training images. Each training image is first warped using a piecewise affine warp from the manually annotated mesh locations in the training images to the base shape. This normalizes each image for shape, so the appearance component of the AAM captures, as far as possible, the changes in the facial features rather than the changes in the images. Thus, the appearance of the AAM is comprised of the pixels that lie inside the base mesh, $\mathbf{x} = (x, y)^{\mathrm{T}} \in \mathbf{s}_0$. PCA is applied (to the shape-normalized images) to provide a compact model of appearance variation of the form:

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{l} \lambda_i A_i(\mathbf{x}) \qquad \forall\ \mathbf{x} \in \mathbf{s}_0, \tag{2}$$

where the coefficients λ_i are the appearance parameters, A_0 is the base appearance, and the appearance images, A_i, are the eigenvectors corresponding to the l largest eigenvalues of the covariance matrix (see Figure 1–b). As with shape, the eigenvectors are the basis vectors that span appearance-space and describe variation in the appearance relative to the mean appearance. The coefficients λ_i are the appearance parameters, which define the contribution of each basis in the reconstruction of $A(\mathbf{x})$. Again, appearance parameters can be considered the coordinates of $A(\mathbf{x})$ in appearance-space, thus each coefficient is a measure of the distance from A_0 to $A(\mathbf{x})$ along the corresponding basis vector.

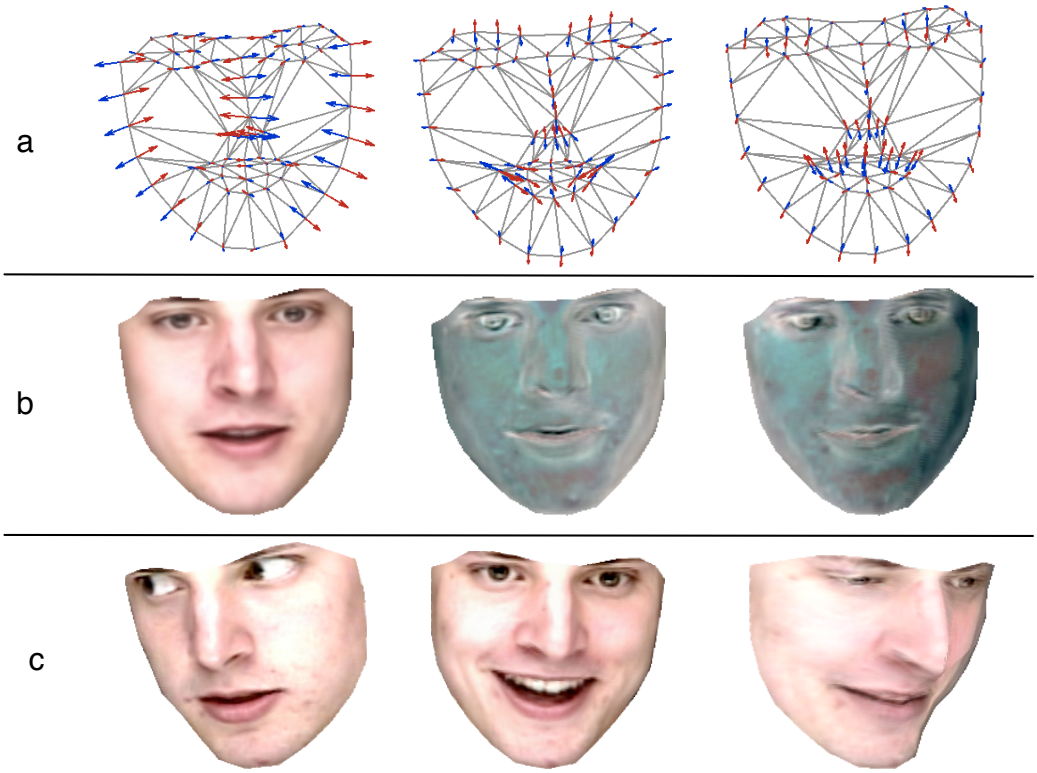

Figure 1. Illustration of Active Appearance Model decomposition of a confederate's face. (a) The first three shape modes of the confederate's AAM shape model. Vectors show movement as each mode score is varied. (b) The mean appearance (left) and first two modes of the confederate's appearance model. (c) Three example facial expressions and poses synthesized using pose coordinates and linear combinations of the first fifteen modes of the AAM illustrated in (a) and (b).

Facial Image Encoding.

To resynthesize an image of the face using an AAM, the shape and appearance parameters that represent the particular face in a particular image are first required. These can be obtained by first annotating the image with the vertices of the AAM shape. Next, given the shape, $\mathbf{s}$, Eq. (1) can be rearranged as follows:

$$p_i = \mathbf{s}_i^T \left(\mathbf{s} - \mathbf{s}_0\right). \tag{3}$$

The image is next warped from $\mathbf{s}$ to $\mathbf{s}_0$ and the appearance parameters computed using:

$$\lambda_i = A_i(\mathbf{x})^T \left(A(\mathbf{x}) - A_0(\mathbf{x})\right). \tag{4}$$

Facial Image Synthesis.

To synthesize an image of the face from a set of AAM parameters, first the shape parameters, $\mathbf{p} = (p_1, \ldots, p_m)^{\mathrm{T}}$, are used to generate the shape, $\mathbf{s}$, of the AAM using Eq. (1). Next the appearance parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_l)^{\mathrm{T}}$ are used to generate the AAM appearance image, $A(\mathbf{x})$, using Eq. (2). Finally a piece-wise affine warp is used to warp $A(\mathbf{x})$ from $\mathbf{s}_0$ to $\mathbf{s}$ (see Figure 1–c).

The advantage of using an AAM rather than the images directly is that the model parameters allow the face to be manipulated prior to resynthesis. For example, we might wish to attenuate or exaggerate the expressiveness of the face, or map facial expressions to different faces. This is difficult to achieve using the images themselves as the raw pixels do not inform directly what the face is doing in an image.

Dissociating Facial Expression and Identity.

To analyse the effects of facial behaviour independently of identity, the two information sources must be separated. AAMs offer a convenient method for doing this efficiently. The shape and appearance basis vectors of an AAM are usually computed with respect to the mean shape and appearance calculated from a set of training images. Given a sufficient number of training examples, in the order of ≈30–50 images, the expression representing the origin of the AAM space will typically converge to the same expression irrespective of the person on which the model is trained — see Figure 2 for example.

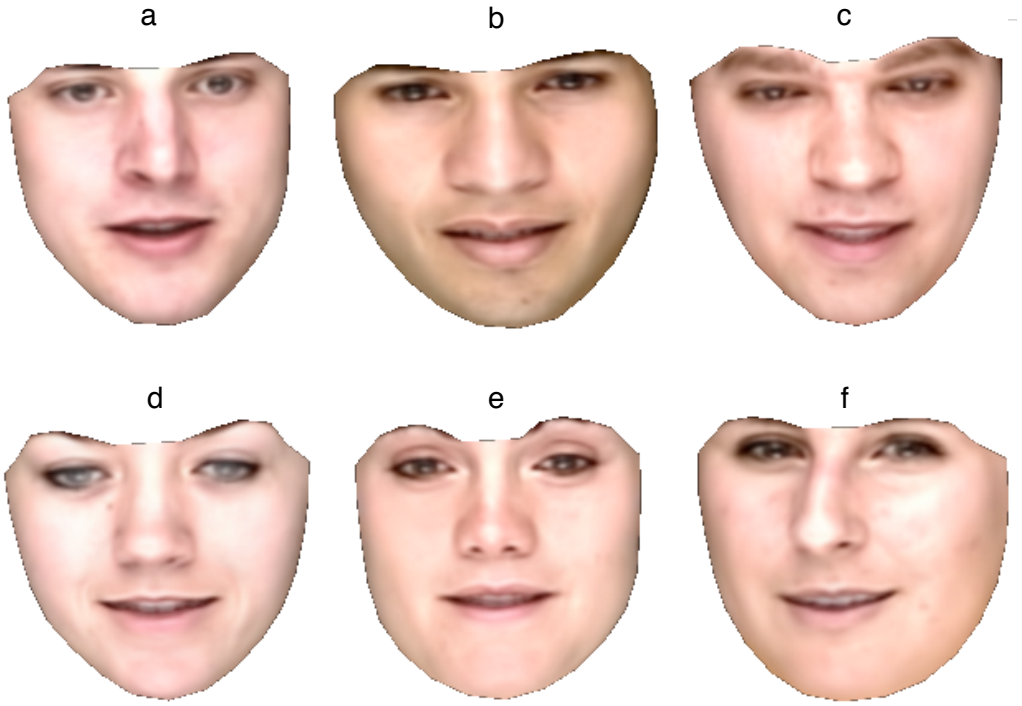

Figure 2. Illustration of the mean shape and appearance for six AAMs each trained on a specific individual converging to approximately the same facial expression. Each model was trained on between 15 and 25 images of the same person.

Thus, the mean shape and appearance of an AAM can be thought of as representative of the identity of the person on which the model was trained. On the other hand, the basis vectors that span the shape and appearance space can be thought of as representative of the changes in facial expression. Thus, to synthesize facial expressions independently of identity, the shape and appearance sub-spaces can be translated by shifting the origin,

which in terms of AAMs involves substituting the mean shape and appearance from one model into another. The effects of this dissociation can be seen in Figure 3 and 4.

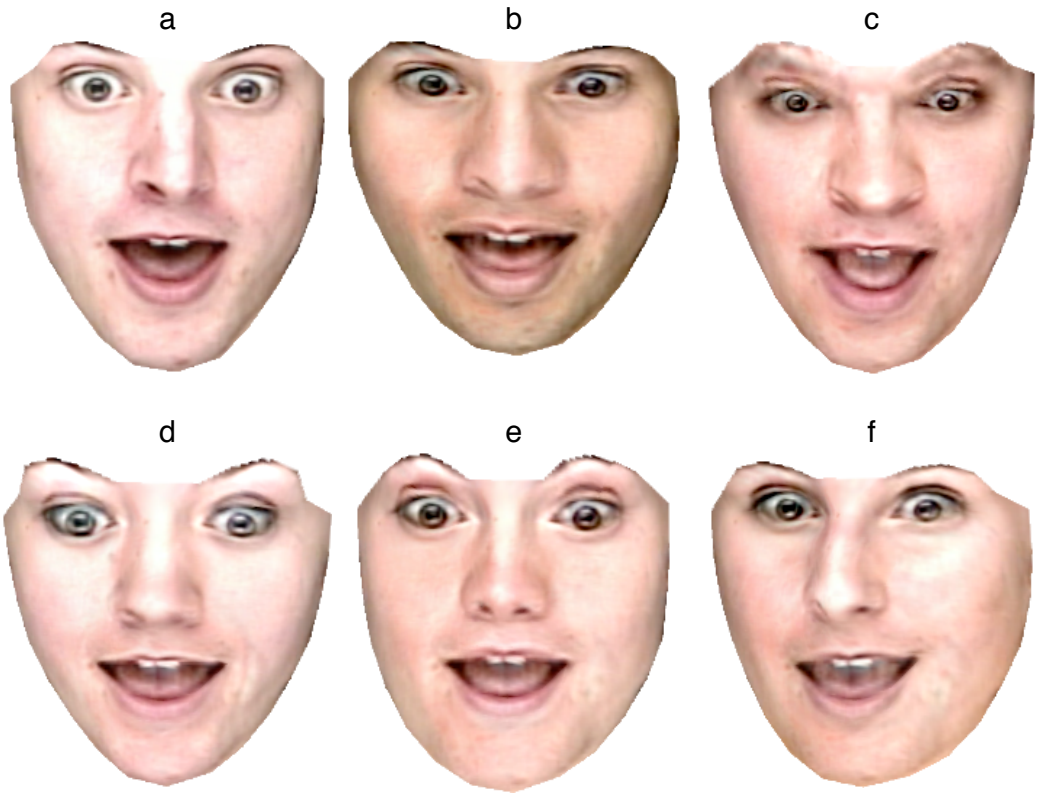

Figure 3. Applying expressions of a male to the appearances of other persons. The top left avatar (a) has the appearance of the person whose motions were tracked. The other columns of the top row (b,c) are avatars with same–sex appearance. The bottom row (d–f) are avatars with opposite–sex appearance.

Apparatus Experiments 1 and 2

Two three–sided 1.2m deep × 1.5m wide × 2.4m tall videoconference booths were constructed of closed cell foam and wood, and set on a 0.5m high wooden stage so as to reduce exposure to ferrous metal that might interfere with the magnetic motion capture. The back of the booth was closed by a black fabric curtain. Each booth was placed in a separate room, but close enough so that a single magnetic field could be used for motion capture of both participants. Color–controlled video lights (2 KinoFlo 4 bulb 1.2m florescent studio light arrays) were placed outside the booth just above and below the 1.5m × 1.2m backprojection screen at the front of the booth. The light arrays had remote ballasts so that the magnetic fields from the ballast transformers could be moved to be more than 4m outside the motion capture magnetic field. White fabric was stretched in front of the video lights and on each side wall in order to diffuse the lighting and reduce shadows.

Motion was recorded by an Ascension Technologies MotionStar system with two Extended Range Transmitters calibrated to emit a synchronized pulsed magnetic field. Each

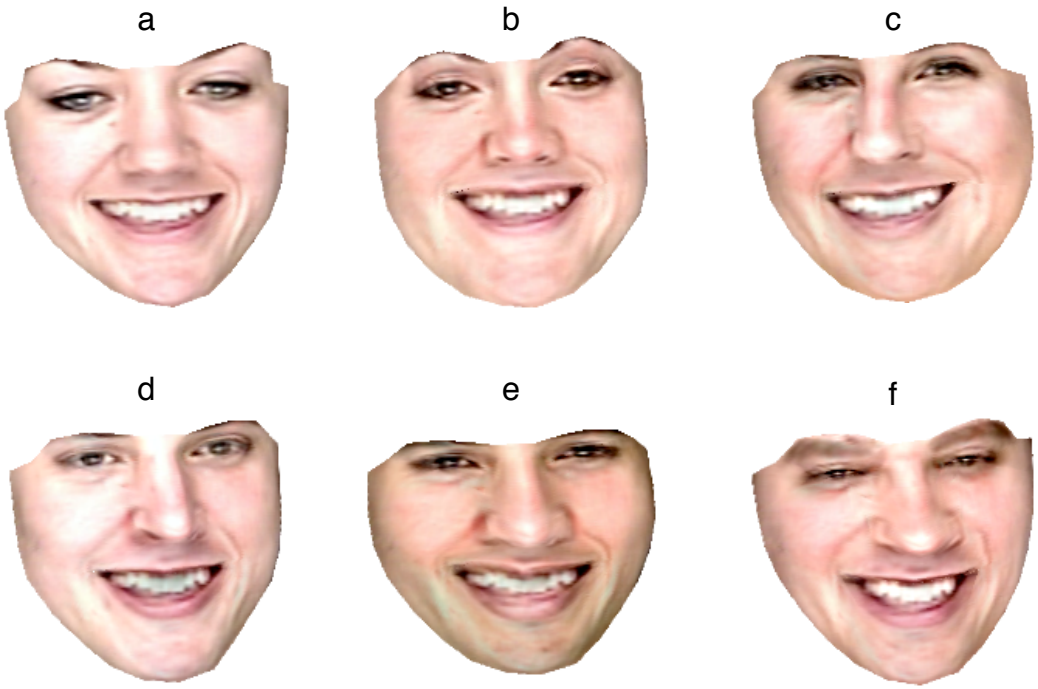

Figure 4. Applying expressions of a woman to the appearances of other persons. (a) Appearance of the person whose motions were tracked. (b–c) Avatars with same–sex appearance. (d–f) Avatars with opposite–sex appearance.

transmitter, a 31cm cube, was located approximately 1.5m from the stool in each booth and outside the black curtain at the back of each booth. The sensors, 2.5cm × 2.5cm × 2.0cm of approximately 16 grams, acted as receivers measuring flux in response to their position and orientation within the transmitter's field. The sensors were attached to the back of each conversants' head using a purpose–made black elastic headband (as shown in Figures 1–a and 1–d). Motion was sampled at 81.6Hz, time–stamped, and stored on an Apple XServe G5.

Video was recorded by a small Toshiba IK-M44H "lipstick camera" positioned just above the head of the image of their interlocutor. Video from the naive participants' booth was routed to a JVC BR-DV600U digital video tape recorder (VTR) and then output to an InFocus X3 video projector for viewing by the confederate. Video from the confederates' booth was routed to another JVC BR-DV600U VTR and then through a Horita VDA-50 distribution amp, splitting the signal into three video streams. The first video stream was sent directly to one input of a Hotronic 8x2 genlocked video switch. The second video stream out of the distribution amp was routed to the input of an AJA Kona card mounted in an Apple quad 2.5GHz G5 PowerMac where AAM modeling was applied. The output of the AJA Kona was then sent to the second input to the video switch. The third video stream was routed through a Hotronic DE41-16RM frame delay (set at two frames delay in order to match the delay induced by the AAM processing and AD/DA of the AJA Kona) and then to a third input of the video switch. Thus, at the genlocked switch, we could alternate

seamlessly between the three possible views of the confederates' booth: undelayed video, 67ms delayed video, or resynthesized avatar. The output of the video switch was then sent to an InFocus X3 video projector for the naive participants' booth. Video images on each back–projection screen were keystone corrected and sized so that the displayed image was life–size. Total video delay from the naive participants' booth to the confederates' booth was 100ms. Total video delay from the confederates' booth to the naive participants' booth was either 100ms or 167ms depending on the experimental condition.

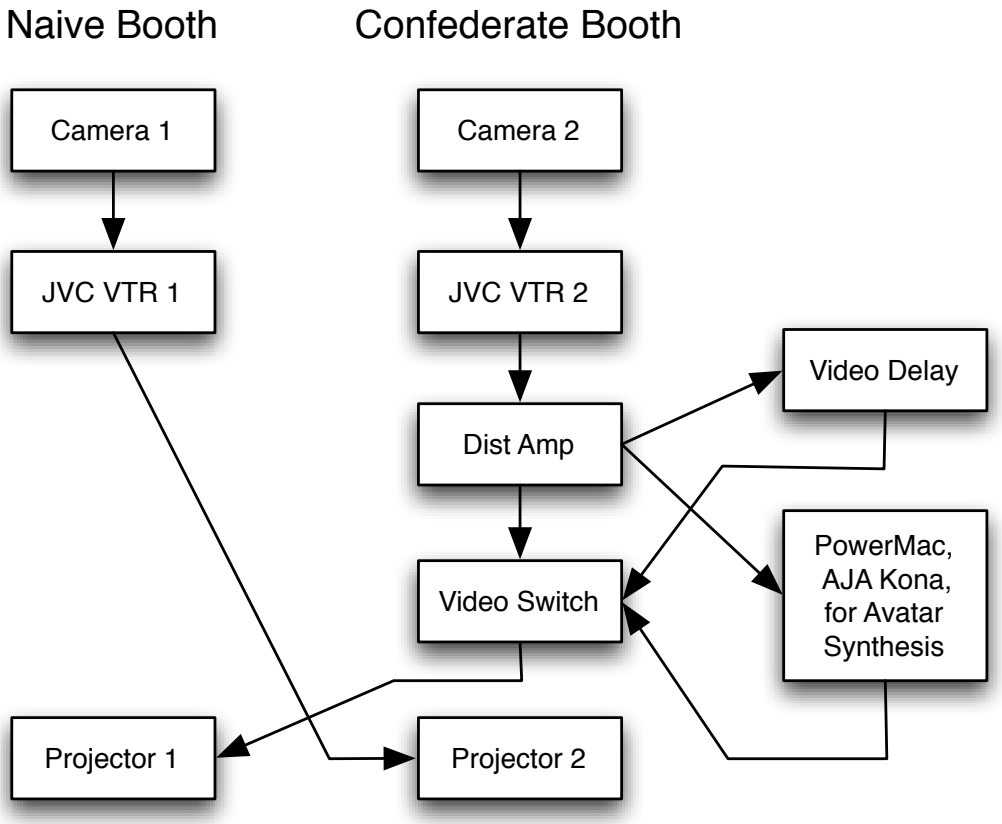

Figure 5. Flowchart of the video display equipment.

Audio was recorded using Earthworks directional microphones outside the field of view of the conversants. The conversants wore lightweight plastic headphones to in order to hear each other. Audio was routed by a Yamaha 01V96 digital mixer which implemented switchable digital delay lines in order to synchronize audio with the total delay induced in each possible video source. The confederates' voices were processed by a TC–Electronics VoicePro pitch and formant processor. When apparent gender was changed in Experiment 2, the pitch of the confederates voices was either raised or lowered and formants changed to be appropriate for someone of the target gender. Audio levels were set to approximate, at the listeners' headphones, the dB level produced by the speaker at the apparent viewing distance.

Experiment 1

In Experiment 1 we verified the believability and measured effects of using the avatar rather than video. Generating the avatar requires 67ms so we tested the effect of this delay. Confederates always saw unprocessed video of the naive participant (as shown in Figure 6–d), but the naive participant saw either either (a) undelayed video, (b) 67ms delayed video, or (c) a resynthsized avatar driven by the motions of the confederate (Figure 6–a or 6–c). We changed the display condition at one minute intervals in counterbalanced order. Confederates were blind to the display order.

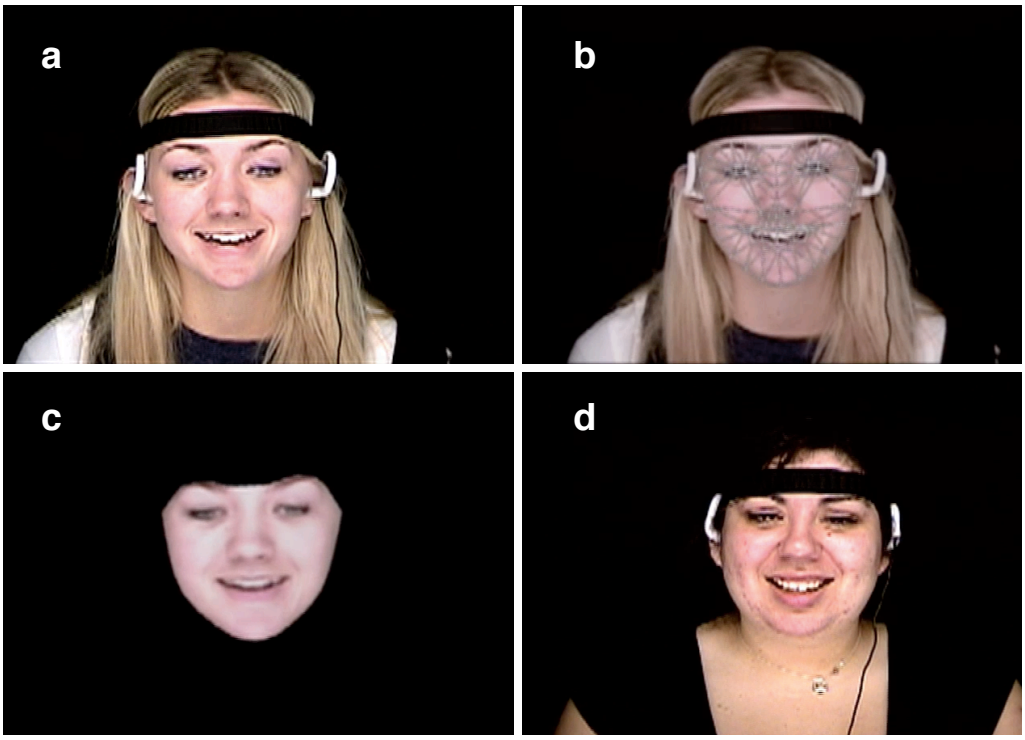

Figure 6. A still frame from Movie S1 recorded during Experiment 2 and included in the supplemental supporting materials. (a) Video of confederate. (b) Model tracking confederate's face. (c) Avatar as it appears to participant. (d) Video of participant.

Methods

Participants.

Confederates ($N = 6$, 3 male, 3 female) were paid undergraduate research assistants who were fully informed of the intent and procedures of the experiment, but were blind to the order of the manipulation during the conversations. Naive participants ($N = 20$, 9 male, 11 female) were recruited from the undergraduate psychology research participant pool and received class credit for participation. All confederates and participants in all experiments reported in the current article read and signed informed consent forms approved by the relevant Institutional Review Board. Images in this article and in the accompanying movie

files were from participants and confederates who signed release forms allowing publication of their images and video.

Procedure.

Confederates and naive participants did not meet other than over the video conference; their video booths were located in separate rooms with separate entries. Confederates were shown video of the naive participant prior to starting the first conversation and if a confederate knew the naive participant, the naive participant was given full credit and dismissed.

Naive participants were informed that they would be engaging in two conversations in a videoconference booth. They were told that we were testing a system that "cut out video to just show the face" and that sometimes they would see the whole person with whom they would be speaking and sometimes they would just see the person's face. We also informed the participants that we would attach a sensor to them and that we were "measuring magnetic fields during conversation." No naive participant guessed that the "cut out" video was a resynthesized avatar or that the sensor measured motion of their head.

Naive participants each engaged in two 16–minute conversations, one with a male confederate and one with a female confederate. At one minute intervals, the video and audio were switched between the three video conditions: undelayed video, 67ms (3 frames) delayed video, and avatar face (which also required a 67ms delay). Since the undelayed video required 100ms to be transmitted to the other booth, the audio was delayed 100ms for the undelayed video and 100ms+67ms=167ms for the delayed video and avatar conditions respectively.

Confederates were instructed to not turn their head more than approximately 20 degrees off axis, since the AAM model would not being able to resynthesize their face due to occlusion. In addition, confederates were instructed to not put their hands to their face during the conversation, since the AAM model would lose tracking of their face.

Confederates always saw the full video and unprocessed audio of the naive participant (as shown in Figure 6–d), but the naive participant saw either video or an avatar (as shown in Figures 6–a and 6–c). After both conversations were finished, naive participants were given a predebriefing questionnaire including the question, "Did anything about the experiment seem odd?" Naive participants were then fully informed about the experiment and asked not to reveal this information to their classmates.

Analyses.

Angles of the head sensor in the anterior–posterior direction and lateral direction were selected for analysis since these directions correspond to the meaningful motion of a head nod and a head turn respectively. We first converted the head angles into angular displacement by subtracting the mean overall head angle across a whole conversation from each head angle sample. We used the overall mean head angle since this provided an estimate of the overall equilibrium head position for each interlocutor independent of the trial conditions. We then low pass filtered the angular displacement time series and calculated angular velocity using a quadratic filtering technique (Generalized Local Linear Approximation, (GLLA) Boker, Deboeck, Edler, & Keel, in press), saving both the estimated displacement and velocity for each sample.

The root mean square (RMS) of the horizontal (RMS–H) and vertical (RMS–V) angular displacement and angular velocity was then calculated for each one minute condition of each conversation for each naive participant and confederate. These measures are equivalent to standard deviations of angular displacement and velocity except that the overall mean displacements rather than the within–trial displacements were used.

We will emphasize the analysis of the angular velocity since this variable can be thought of as how animated a participant was during an interval of time. The angular displacement is a similar variable, measuring how far a person was, on average, from their mean head pose during a trial. But RMS angular displacement could be high either because a participant was moving in an animated fashion, or could be high because, within a trial, the participant was looking away from her or his mean head pose. Thus both RMS displacement and velocity are informative, but RMS velocity is a measure of what is most often thought of as head movement.

We expect that the naive participant affected the head movements of the confederate as well as the confederate affecting the head movements of the participant. We might expect that the video manipulations have an effect on the naive participant, who has an effect on the confederate, who in turn has an effect back on the naive participant. Given these bi–directional feedback effects, these data need to be analyzed while taking into account both participants in the conversation simultaneously. Each interlocutor's head movements are thus both a predictor variable and outcome variable. Neither can be considered to be an independent variable. In addition, each naive participant was engaged in two conversations, one with each of two confederates. Each of these sources of non–independence in dyadic data need to be accounted for in a statistical analysis (Kenny & Judd, 1986).

In order to put both interlocutors in a dyad into the same analysis we used a variant of Actor–Partner analysis (Kashy & Kenny, 2000; Kenny, Mannetti, Pierro, Livi, & Kashy, 2002). Suppose we are analyzing RMS–V angular velocity. We place both the naive participants' and confederates' RMS–V angular velocity into the same column in the data matrix and use a second column as a dummy code labeled "Confederate" to identify whether the data in the angular velocity column came from a naive participant or a confederate. In a third column, we place the RMS–V angular velocity from the other participant in the conversation. We then use the terminology "Actor" and "Partner" to distinguish which variable is the predictor and which is the outcome for a selected row in the data matrix. If Confederate=1, then the confederate is the "Actor" and the naive participant is the "Partner" in that row of the data matrix. If Confederate=0, then the naive participant is the "Actor" and the confederate is the "Partner".

We then coded the sex of the "Actor" and the "Partner" as a binary variables (0=female, 1=male). The RMS angular displacement of the "Partner" was used as a continuous predictor variable. Binary variables were coded for each manipulated condition: delay condition (0=100ms, 1=167ms) and avatar condition (0=video, 1=avatar). Since only the naive participant sees the manipulated conditions we also added two interaction variables (delay condition $\times$ confederate and avatar $\times$ confederate), centering each binary variable prior to multiplying. The manipulated condition may effect the naive participant directly, but also may effect the confederate indirectly through changes in behavior of the naive participant. The interaction variables allow us to account for an overall effect of the manipulation as well as differences between the naive participant and confederate.

We then fit models using restricted maximum likelihood using the R function lme(). Since there is non–independence of rows in this data matrix, we need to account for this non–independence. An additional column is added to the data matrix that is coded by experimental session and then the mixed effects model of the data is grouped by experimental session column (both conversations in which the naive participant engaged). Each session was allowed a random intercept to account for individual differences between experimental sessions in the overall displacement or velocity. This model can be expressed as a mixed effects model

$$\begin{aligned} y_{ij} &= b_{j0} + b_1 SA_{ij} + b_2 SP_{ij} + b_3 C_{ij} + b_4 A_{ij} + b_5 D_{ij} + \\ &\quad b_6 PY_{ij} + b_7 A_{ij} C_{ij} + b_8 D_{ij} C_{ij} + e_{ij} \qquad (5) \\ b_{j0} &= c_{00} + u_{j0} \qquad (6) \end{aligned}$$

where y_{ij} is the outcome variable (either RMS–H or RMS–V angular displacement or velocity) for condition i and session j. The other predictor variables are the sex of the Actor SA_{ij}, the sex of the Partner SP_{ij}, whether the Actor is the confederate C_{ij}, the avatar display condition A_{ij}, the delay display condition D_{ij}, the RMS–H or RMS–V of the partner PY_{ij}, and the two interactions $A_{ij} \times C_{ij}$ and $D_{ij} \times C_{ij}$. Since each session was allowed to have its own intercept, the predictions are relative to the overall angular displacement and velocity associated with each naive participant.

Results

Prior to debriefing, we asked participants, "Did anything about the experiment seem odd?". Even though the participants were being shown alternate views of the video and the avatar, thereby maximizing contrast effects, no participant responded that they thought the avatar face was computer–generated or expressed doubt about our cover story that the faces with which they spoke were "video cut out around the face".

Table 1 presents the results of the mixed effects random intercept model grouped by session (the two dyadic conversations in which each naive participant engaged). The intercept, $10.417°$/sec ($p < 0.0001$) is the overall mean vertical RMS angular velocity of both confederates and naive participants. The line "Actor is Male" displays a point estimate of $-1.202°$/sec ($p < 0.0001$), thus males' vertical angular velocity was estimated to have about one degree per second smaller RMS–V angular velocity than females. The next line, "Partner is Male" estimates that when the conversational partner was male, the person being measured displayed about one degree per second ($p < 0.001$) smaller RMS–V angular velocity than when the interlocutor was female. The third line, "Actor is Confederate" suggests that the confederates produced RMS–V angular velocity that was not significantly different from the naive participants. The next two lines suggest that neither the avatar nor the delay produced a significant effect on the RMS–V angular velocity.

The "Partner Vertical RMS" line estimates that there was a reciprocal relationship in the two interlocutors' RMS–V angular velocity such that for each $1°$/sec increase in the interlocutor's RMS–V there was a $0.164°$/sec decrease in the RMS–V angular velocity of the person being measured ($p < 0.001$). Finally, the interaction terms were not significantly different from zero.

The RMS–V angular displacement results are shown in Table 2. The significant effects have the same pattern as for RMS–V angular velocity except that there is a significant main

Table 1: Head vertical RMS angular velocity in degrees from Experiment 1 predicted using a mixed effects random intercept model grouped by session. Observations= 530, Groups=20, AIC=2600.7, BIC=2647.5. The standard deviation of the random intercept was 1.634.

	Value	SE	DF	t	p
Intercept	10.417	0.5982	502	17.41	$< .0001$
Actor is Male	-1.202	0.2787	502	-4.31	$< .0001$
Partner is Male	-1.042	0.2800	502	-3.72	0.0002
Actor is Confederate	-0.312	0.2303	502	-1.36	0.1759
Avatar Display	-0.038	0.3016	502	-0.13	0.8999
Delayed Display	0.111	0.2763	502	0.40	0.6890
Partner Vertical RMS	-0.164	0.0432	502	-3.80	0.0002
Confederate × Avatar	0.172	0.6031	502	0.28	0.7762
Confederate × Delay	0.261	0.5523	502	0.47	0.6369

effect of the confederate such that confederates exhibited lower RMS–V angular displacement than naive participants. Note that the coefficients for "Actor is Male" and "Partner is Male" are within two standard errors of these coefficients reported for face–to–face dyadic conversations (Ashenfelter et al., in press).

Table 2: Head vertical RMS angular displacement in degrees from Experiment 1 predicted using a mixed effects random intercept model grouped by session. Observations= 530, Groups=20, AIC=1723.1, BIC=1769.9. The standard deviation of the random intercept was 0.924.

	Value	SE	DF	t	p
Intercept	4.811	0.2761	502	17.43	$< .0001$
Actor is Male	-0.592	0.1205	502	-4.92	$< .0001$
Partner is Male	-0.508	0.1212	502	-4.19	$< .0001$
Actor is Confederate	-0.806	0.1076	502	-7.49	$< .0001$
Avatar Display	0.067	0.1289	502	0.52	0.6043
Delayed Display	-0.001	0.1181	502	-0.01	0.9926
Partner Vertical RMS	-0.216	0.0428	502	-5.05	$< .0001$
Confederate × Avatar	-0.049	0.2578	502	-0.19	0.8489
Confederate × Delay	0.029	0.2360	502	0.12	0.9028

Results for the mixed effects model applied to RMS–H angular velocity are presented in Table 3. Men exhibited 90.619°/sec less RMS–H angular velocity than women ($p < 0.0001$) and 41.263°/sec less RMS–H ($p < 0.01$) was exhibited when the conversational partner was male rather than female. Confederates exhibited 70.333°/sec less ($p < 0.0001$) RMS–H than naive participants. No effects were found for either the avatar or delayed display conditions or for their interactions with the Confederate. There was a compensatory effect such that when one conversant's RMS–H was 1.0°/sec greater, the other conversant's RMS–H was 0.361°/sec less ($p < 0.0001$).

Table 3: Head horizontal RMS angular velocity in degrees from Experiment 1 predicted using a mixed effects random intercept model grouped by session.Observations= 530, Groups=20, AIC=6781.2, BIC=6828.0. The standard deviation of the random intercept was 134.9.

	Value	SE	DF	t	p
Intercept	320.539	34.708	502	9.24	$< .0001$
Actor is Male	-90.619	15.145	502	-5.98	$< .0001$
Partner is Male	-41.263	15.552	502	-2.65	0.0082
Actor is Confederate	-70.333	13.204	502	-5.33	$< .0001$
Avatar Display	15.242	16.306	502	0.93	0.3504
Delayed Display	1.744	14.933	502	0.12	0.9071
Partner Horizontal RMS	-0.361	0.041	502	-8.83	$< .0001$
Confederate × Avatar	4.472	32.593	502	0.14	0.8909
Confederate × Delay	8.070	29.845	502	0.27	0.7870

Results for the mixed effects model applied to RMS–H angular displacement are presented in Table 4. The pattern of significant effects and the signs of their coefficients is identical to that for RMS–H angular velocity.

Table 4: Head horizontal RMS angular displacement in degrees from Experiment 1 predicted using a mixed effects random intercept model grouped by session. Observations= 530, Groups=20, AIC=5285.4, BIC=5332.2. The standard deviation of the random intercept was 31.52.

	Value	SE	DF	t	p
Intercept	80.548	8.1513	502	9.88	$< .0001$
Actor is Male	-22.143	3.6141	502	-6.13	$< .0001$
Partner is Male	-9.082	3.7209	502	-2.44	0.0150
Actor is Confederate	-19.157	3.2114	502	-5.97	$< .0001$
Avatar Display	2.336	3.8934	502	0.60	0.5488
Delayed Display	0.756	3.5665	502	0.21	0.8321
Partner Horizontal RMS	-0.367	0.0408	502	-9.01	$< .0001$
Confederate × Avatar	0.086	7.7840	502	0.01	0.9912
Confederate × Delay	2.937	7.1295	502	0.41	0.6806

Discussion

The primary result of Experiment 1 is that there were strong effects for all predictors except Avatar Display and Delayed Display and their interactions with the Confederate variable. Thus, a naive participant or confederate did not move his or her head any more or less when the avatar was displayed or when the display was delayed. In addition naive participants were not significantly different in their reactions to the manipulated conditions than confederates. This result is consistent with the fact that even though we switched between raw video and the avatar, no participant expressed doubts about the cover story

that the avatar displays were "video cut out around the face" when given an opportunity to do so prior to debriefing.

Experiment 1 exhibits sex effects for actor and partner such that when the actor or partner is a male, there is less horizontal and vertical RMS angular velocity and angular displacement. Women move their heads more actively and with greater extent than men in this experiment and also, when speaking with a woman, both men and women tend to move their heads more actively and with greater extent than when speaking with a man. This finding replicates that of Ashenfelter et al. (in press) who motion tracked head movements in naive participants in face–to–face dyadic conversations. For RMS–V, parameter values were very similar between these two experiments. However, for RMS–H, the videoconference paradigm in the current experiment is substantially larger than was found in the face–to–face paradigm. It may be that some of this difference is attributable to differences between how people behave in face–to–face conversations relative to videoconferences. It may be that in videoconferences people feel less co–presence and thus orient themselves less often directly towards their conversational partner. This may also have to do with cues to break symmetry (Boker & Rotondo, 2002; Ashenfelter et al., in press), in other words it may take a larger movement in a videoconference to produce a cue that behavioral mirroring with one's partner should no longer be in effect.

Finally, there was evidence for a reciprocal effect of head movements in both the horizontal and vertical direction. Thus over the one minute trials, greater than normal movement by one conversant is associated with less than normal movement by the other conversant. This effect could be partially due to imbalance in speaker–listener roles during the trials or could be also be attributed to symmetry breaking. In either case, it is evidence of negative coupling between the amplitude of conversants' head movements.

Experiment 2

In Experiment 2 we substituted the appearance of each confederate for every other confederate as illustrated in Figures 3 and 4. A confederate's resynthesized avatar might be him– or herself, or someone of the same or opposite sex. Confederates knew of the manipulation, but were blind to what identity and sex they appeared to be in any given conversation.

Methods

Participants.

Confederates ($N = 6$, 3 male, 3 female) were the same confederates as in Experiment 1. Naive participants ($N = 28$, 11 male, 17 female) were recruited from the undergraduate psychology research participant pool and received class credit for participation. One participant appeared to know about the experiment in advance and was given full credit and dropped from the experiment.

Procedure.

Confederates and naive participants did not meet other than over the video conference; their video booths were located in separate rooms with separate entries. Confederates were shown video of the naive participant prior to starting the first conversation and if

a confederate knew the naive participant, the naive participant was given full credit and dismissed.

Naive participants were informed that they would be engaging in six conversations in a videoconference booth. They were told that we "cut out video to just show the face so that you only pay attention to the other person's face". We also informed the participants that we would attach a sensor to them and that we were "measuring magnetic fields during conversation." Only one naive participant guessed that the "cut out video" was a resynthesized avatar or that the sensor measured motion of their head. It was evident from the beginning of this participant's first conversation that he had been previously informed of the nature of the experiment. He was given credit and sent away without completing the experiment.

Naive participants each engaged in six 4–minute conversations, three conversations with one male confederate and three conversations with one female confederate. At the beginning of each of the six conversations, the two conversants were given a topic to discuss: Classes/Major/Career, Movies and TV, Weekend, Travel, Music, or a local sports team.

Each confederate appeared as three different avatars, an avatar constructed from the model for him or herself, an avatar of another confederate of the same sex, or an avatar of another confederate of the opposite sex. The confederates were blind to which avatar the naive participant saw during each of their three conversations. Thus, the naive participant appeared to have conversations with six different individuals, although in actuality they only spoke with two individuals.

Confederates were again instructed to not turn their head more than approximately 20 degrees off axis and to not put their hands to their face during the conversation. In addition, the confederates were instructed to restrict their dialog so as to not give away what sex they were by revealing information such as their name, what dorm they lived in, gender of roommates, etc.

Confederates always saw the full video and unprocessed audio of the naive participant (as shown in Figure 6–d), but the naive participant saw an avatar (as shown in Figures 6–a and 6–c) and heard processed audio that was pitch and formant shifted to be appropriate to the apparent sex of the avatar. After both conversations were finished, naive participants were given a predebriefing questionnaire including the question, "Did anything about the experiment seem odd?" No participant reported that they thought that they were speaking with a computer animation or that they were speaking with fewer than 6 people. Naive participants were then fully informed about the experiment and asked not to reveal this information to their classmates.

To transform between confederates' appearances, we trained avatar models from video of each confederate performing facial expressions characteristic of all other confederates. From recordings of the confederates engaged in conversation in Experiment 1, we created a video of each confederate making characteristic utterances, expressions, and movements. Each confederate was then videotaped while imitating this video of all confederates' mannerisms. Then, the resulting video was used to train models that preserved a mapping from each confederate to every other confederate. Finally, we created pitch and formant shifting programs for a TC–Electronic VoiceOne vocal processor to transform each source voice as much as possible into each target voice.

Analyses.

RMS vertical and horizontal angular velocity and displacement were calculated in the same manner as in Experiment 1. Predictor variables unique to Experiment 2 were the sex of the avatar (0=female, 1=male), whether the avatar displayed was the confederate's avatar (0=other face, 1=same face), and whether the avatar was of same or opposite sex to the confederate (0=same sex, 1=opposite sex).

RMS–H and RMS–V angular velocity and displacement were predicted using mixed effects random intercept models grouped by session,

$$\begin{aligned} y_{ij} &= b_{j0} + b_1 SA_{ij} + b_2 SP_{ij} + b_3 C_{ij} + b_4 A_{ij} + b_5 D_{ij} + \\ &\quad b_6 PY_{ij} + b_7 A_{ij} C_{ij} + b_8 D_{ij} C_{ij} + e_{ij} \qquad (7) \\ b_{j0} &= c_{00} + u_{j0} \qquad (8) \end{aligned}$$

where y_{ij} is the outcome variable (either RMS–H or RMS–V angular displacement or velocity) for condition i and session j. The other predictor variables are the sex of the Actor SA_{ij}, the sex of the Partner SP_{ij}, whether the Actor is the confederate C_{ij}, the sex of the avatar display A_{ij}, the RMS–H or RMS–V of the partner PY_{ij}, and the two interactions $A_{ij} \times C_{ij}$ and $SP_{ij} \times C_{ij}$.

Results

The results from the mixed effects random intercept model of RMS–V angular velocity are displayed in Table 5. In this model, the effects of actual sex of the participants and the apparent sex of the avatar are simultaneously estimated. The "Actor is Male" and "Partner is Male" effects are the same sign and more than two standard errors larger than the equivalent effects in Experiment 1. On the other hand, the coefficient for the sex of the Avatar is not significantly different from zero. Thus the effect of the sex of the interlocutor on RMS–V angular velocity is primarily due to the motions and vocal prosody of the interlocutor and not due to the facial appearance and pitch of the voice.

There is a reciprocal effect such that during a 60s segment when one person's RMS–V angular velocity is 1°/sec larger, the other person's RMS–V is 0.484°/sec smaller ($p < 0.0001$). This negative coupling effect is more than two standard errors stronger than in Experiment 1.

The mixed effects model results for RMS–V angular displacement are displayed in Table 6. The pattern of significant results coincides with that of RMS–V angular velocity. The "Actor is Male" and "Partner is Male" effects are the same sign and well within 1 standard error of the equivalent effects in Experiment 1.

Table 7 presents the results of the mixed effects model for RMS–H angular velocity. The pattern of significant results and the signs of their coefficients are identical to that of RMS–V angular velocity, although the "Partner is Male" effect only barely achieves statistical significance ($p < 0.05$).

When RMS–H angular displacement was predicted from the same variables (see Table 8), there were only two significant effects. First, the confederates turned their heads less than the naive participants ($p < 0.0001$). Also, there was a reciprocal horizontal effect such that during a 60s segment when one person's RMS–H was 1° larger the other person's RMS–H was 0.40° smaller ($p < 0.0001$).

Table 5: Head RMS–V angular velocity from Experiment 2 predicted using a mixed effects random intercept model grouped by session. Observations=310, Groups=28, AIC=1043.2, BIC=1080.3. The standard deviation of the random intercept was 1.527.

	Value	SE	DF	t	p
Intercept	13.324	0.5261	275	25.325	$< .0001$
Actor is Male	-2.769	0.2385	275	-11.608	$< .0001$
Partner is Male	-1.772	0.2606	275	-6.801	$< .0001$
Actor is Confederate	-0.872	0.1924	275	-4.532	$< .0001$
Avatar is Male	-0.070	0.1820	275	-0.384	0.7016
Partner Vertical RMS	-0.484	0.0506	275	-9.564	$< .0001$
Confederate × Avatar Sex	0.008	0.3598	275	0.023	0.9814
Confederate × Partner Sex	-0.257	0.4720	275	-0.545	0.5859

Table 6: Head RMS–V angular displacement from Experiment 2 predicted using a mixed effects random intercept model grouped by session. Observations=310, Groups=28, AIC=1043.2, BIC=1080.3. The standard deviation of the random intercept was 0.796.

	Value	SE	DF	t	p
Intercept	4.663	0.2646	275	17.623	$< .0001$
Actor is Male	-0.448	0.1738	275	-2.576	0.0105
Partner is Male	-0.537	0.1712	275	-3.139	0.0019
Actor is Confederate	-1.222	0.1558	275	-7.838	$< .0001$
Avatar is Male	0.169	0.1395	275	1.214	0.2259
Partner Vertical RMS	-0.153	0.0571	275	-2.677	0.0079
Confederate × Avatar Sex	-0.198	0.2754	275	-0.718	0.4734
Confederate × Partner Sex	-0.275	0.3431	275	-0.801	0.4237

Table 7: Head RMS–H angular velocity from Experiment 2 predicted using a mixed effects random intercept model grouped by session. Observations=310, Groups=28, AIC=3824.8, BIC=3861.9. The standard deviation of the random intercept was 105.22.

	Value	SE	DF	t	p
Intercept	261.74	25.593	275	10.227	$< .0001$
Actor is Male	-50.61	17.422	275	-2.905	0.0040
Partner is Male	-34.36	17.182	275	-2.000	0.0465
Actor is Confederate	-56.19	13.736	275	-4.091	0.0001
Avatar is Male	-19.06	13.424	275	-1.420	0.1567
Partner Horizontal RMS	-0.44	0.052	275	-8.409	$< .0001$
Confederate × Avatar Sex	29.09	26.682	275	1.090	0.2765
Confederate × Partner Sex	-46.32	34.771	275	-1.332	0.1839

Table 8: Head RMS–H angular displacement from Experiment 2 predicted using a mixed effects random intercept model grouped by session. Observations=310, Groups=28, AIC=3012.1, BIC=3049.2. The standard deviation of the random intercept was 20.72.

	Value	SE	DF	t	p
Intercept	64.99	5.736	275	11.330	$< .0001$
Actor is Male	-5.24	4.467	275	-1.172	0.2422
Partner is Male	-0.94	4.403	275	-0.212	0.8320
Actor is Confederate	-16.06	3.688	275	-4.355	$< .0001$
Avatar is Male	-3.65	3.586	275	-1.017	0.3098
Partner Horizontal RMS	-0.40	0.053	275	-7.595	$< .0001$
Confederate × Avatar Sex	8.20	7.118	275	1.152	0.2503
Confederate × Partner Sex	-11.11	8.845	275	-1.256	0.2102

Discussion

The manipulation of identity and sex was convincing in that no naive participant expressed doubts about the cover story when given an opportunity to do so prior to full debriefing. One participant was dropped because he could not be considered to be naive since he had apparently been informed of the manipulations used in the experiment prior to his beginning of the experiment. The debriefing informed the naive participants that although they may have assumed that they spoke with six different persons, in fact the images they saw were computer–synthesized avatars and that the first three conversations were with the same person and the second three were with a different person. Several participants were convinced that the debriefing itself was the deception in the experiment and refused to believe that they had only spoken with two confederates and that sometimes the sex of the avatar did not match the confederate. Identity information has been reported to be strongly connected with the perceived form of the face relative to lesser contributions from rigid and nonrigid motion cues (Knappmeyer et al., 2003), and this relative bias towards perception of identity from form is likely to have been a contributor to participants' belief that they had been speaking with six different individuals.

For RMS–V displacement and velocity, Tables 1 and 5 and Tables 2 and 6 are strikingly similar in all of the coefficients that are the estimated from the same predictor variables across the two experiments. On the other hand, we found no effects for the sex of the avatar in either velocity or displacement of the vertical head angles. Thus, the apparent gender of the person one is speaking with appears not to affect the amplitude or velocity of one's vertical angular head movement.

In both Experiment 1 and 2 we find overall negative coupling such that during one minute segments, greater RMS–V angular velocity or displacement in one conversant is associated with less RMS–V in the other conversant. This negative coupling has the effect of forcing the two conversants' vertical head movements away from the common equilibrium.

For RMS–H velocity, Tables 3 and 7 again show the same pattern of significant effects and their signs for equivalent coefficients. Also, as in the RMS–V velocity and displacement, there is no effect of the apparent sex of the avatar.

Finally, in RMS–H displacement there is again no effect of the apparent sex of the avatar. However, there is a difference between Tables 4 and 8 in that there are sex effects for both actor and partner in Experiment 1, but no sex effects for either the actor or partner in Experiment 2. This suggests that there is something different about the two experiments, but only in the angular extent of head turning and not in head turning velocity. In a previous face–to–face experiment (Ashenfelter et al., in press), the Actor is Male effect was found to be similar to that found in Experiment 1 for both RMS–V and RMS–H. We do not yet have a good explanation for this puzzling lack of sex effects in RMS–H angular displacement in Experiment 2.

Experiment 3

A possible threat to the results from Experiment 2 is that apparent sex of the avatar might not have been perceived as intended by the manipulation. In part to check this possibility, we performed a rating experiment where raters were asked to view short (10s) video clips of the avatars and select the sex of the person shown in the video. At the same time, we wished to ascertain to what degree ratings of masculinity and femininity are influenced by appearance relative being influenced by dynamics. A hypothesis of separate perceptual streams for appearance and biological motion would suggest that there could be independent contributions from appearance and dynamics to these ratings. Given the capabilities of our software to be able to randomly assign appearance, we were able to construct stimuli so that appearance was counterbalanced: each video clip was shown with a male appearance to half the raters and with a female appearance to the other half of the raters.

Participants.

Participants were 81 men and women from a city several hundred miles from where Experiments 1 and 2 were run. Data from nine participants were excluded because of failure to follow directions or incomplete data. A total of 72 participants (39 male, 33 female) were included in the analysis. Participants' age ranges from 18 to 40 years old, with an average of 22.1 years old.

Procedure.

Participants were randomly assigned into 1 of 6 groups. Each group watched a series of 48 items which consist of three types of presentations (video+audio, video only, and still image+audio). The order of the items were determined at random. Items were projected one at a time onto a large viewing screen to groups of 2–6 participants. Participants recorded their judgments during a pause following each item. They were instructed to watch the whole clip and make judgments after seeing the item number at the end of the clip. Participants made a categorical judgment of gender (male or female) and Likert–type ratings of femininenes and masculineness (from 1 = not at all, to 7 = extremely). The order of femininenes and masculineness ratings was counter balanced so that half of the subjects rated femininenes first and the other half rated masculineness first.

Analyses.

Multiple judgments of sex and ratings of masculinity and femininity were given by each of the raters, so a mixed effects model random coefficients model grouped by rater

was selected for the analyses. Since sex of the rater may have had an effect on his or her ratings, we used sex of the rater (0=female, 1=male) as a predictor at the second level. Fitting this model is equivalent to a level one model where sex of the rater is an interaction term. Predictor variables were sex of the avatar, sex of the confederate, and sex of the rater (0=female, 1=male), whether the sex of the avatar and confederate matched (0=opposite sex, 1=same sex), and whether the displayed movie clip included audio (0=silent, 1=with audio), and whether the displayed clip included motion (0=still, 1=motion). In order to avoid spurious correlations induced by the interaction terms, we centered all predictor variables by subtracting their means prior to fitting each model. The results of fitting the three models are displayed in Tables 9, 10, and 11.

We randomly selected 48 10–second clips (8 from each of the 6 confederates) from video recorded during Experiment 2. We re–rendered each clip once as a male and once as a female, neither avatar being the same individual as the source video. Each of the 48 clips was shown either with motion+audio, motion+silence, or still+audio to viewers blocked so that each viewer saw each source video clip only once and saw each source confederate only rendered as either male or female, but not both. All raters (N = 72, 39 men, 33 women) read and signed IRB–approved informed consent.

Results

Raters judged the sex of the clip consistent with the sex of the avatar 91.9% of the time. To give an idea of what conditions led to judgments of sex that did not agree with the sex of the avatar, we disaggregated by sex of the confederate and sex of the avatar. The majority of cases where there was disagreement between raters judgment of sex and the sex of the avatar occurred when male confederates were displayed as female avatars: In that case, the clips were judged to be male 18.7% of the time. Female confederates displayed as male avatars were judged to be female only 6.1% of the time. But, there were mismatches even when the sex of the confederate and sex of the avatar were consistent. Male confederates displayed as male avatars were judged to be female 4.6% of the time. Female confederates displayed as female avatars were judged to be male 2.8% of the time.

Raters' judgments of sex were also predicted by a mixed effects random intercept model grouped by rater. Since raters were given a forced binary choice as the outcome variable (coded male=1, female=0), a binomial link function was used and fixed parameter estimates are displayed in Table 9. The largest effect is the sex of the avatar (coded male=1, female=0). This effect is positive and so raters judged the sex of the person displayed in the clip to be the same as the apparent sex of the avatar. There was also a smaller, but statistically significant, effect such that when the clip included audio, raters were more likely to choose female. There are also two significant interactions such that when audio is present, the effects of the avatar's sex as well as the confederate's sex are increased. Since the main effect of the sex of the confederate is not significant, it appears that only when audio is present does the sex of the confederate influence the raters' judgments of the sex of the displayed face.

Results of modeling ratings of femininity on a 7 point scale are displayed in Table 10. The strongest effect is the sex of the avatar, where displaying a female avatar is associated with a 2.0 point difference in femininity rating. There is also a main effect of the sex of the confederate, where a female confederate is associated with a 0.5 point difference in femininity

Table 9: Judgments of avatar sex from Experiment 3 predicted by a mixed effects random intercept model with a binomial link function grouped by rater. (Observations: 3454, Groups: 72, AIC= 1700, BIC=1768)

	Value	SE	z	p
Sex Intercept	-2.4685	0.58694	-4.206	$< .0001$
Male Avatar	4.6218	0.65175	7.091	$< .0001$
Male Confederate	0.2036	0.60033	0.339	0.7345
Has Audio	-1.4341	0.44506	-3.222	0.0013
Has Motion	-0.2964	0.53486	-0.554	0.5794
Male Avatar × Has Audio	2.2737	0.49946	4.552	$< .0001$
Male Avatar × Has Motion	0.2904	0.61479	0.472	0.6367
Male Confederate × Has Audio	2.2928	0.46959	4.883	$< .0001$
Male Confederate × Has Motion	0.0712	0.56197	0.127	0.8992
Male Rater	0.1434	0.21206	0.676	0.4990

rating. There are two other main effects for the display methods: displays with audio are associated with a 0.4 lower femininity rating and displays with motion are associated with 0.2 increase in femininity ratings. In addition, there are two significant interactions with audio. When a female confederate is displayed with audio, there is an addition 1.0 point gain in the femininity rating, and when a female avatar is displayed with audio, there is a 0.4 lower femininity rating.

Table 10: Ratings of avatar femininity (7 point Likert scale) from Experiment 3 predicted by a mixed effects random intercept model grouped by rater with sex of rater as a second level predictor.(Observations: 3456, Groups: 72, AIC= 11710, BIC=11783)

	Value	SE	DF	t	p
Femininity Intercept	5.0801	0.13933	3376	36.460	$< .0001$
Female Avatar	2.0365	0.13081	3376	15.568	$< .0001$
Female Confederate	0.4670	0.13081	3376	3.570	0.0004
Female Rater	0.0076	0.11413	70	0.067	0.9469
Has Audio	-0.4089	0.09250	3376	-4.420	$< .0001$
Has Motion	0.2431	0.09250	3376	2.628	0.0086
Female Avatar × Has Audio	-0.3628	0.10680	3376	-3.397	0.0007
Female Avatar × Has Motion	0.0729	0.10680	3376	0.683	0.4948
Female Confederate × Has Audio	1.0503	0.10680	3376	9.834	$< .0001$
Female Confederate × Has Motion	-0.0521	0.10680	3376	-0.488	0.6258

The model of masculinity (Table 11) exhibits similar effects for sex of avatar (2.1 point masculinity increase for male avatars), but there is a main effect for sex of confederate (0.3 point masculinity increase for male confederates) and no main effect for audio or motion. However, there are two significant interactions with audio: male confederates displayed

with audio are associated with a 0.9 increase in masculinity and male avatars displayed with audio are associated with a 0.3 point decrease in masculinity ratings.

Table 11: Ratings of avatar masculinity (7 point Likert scale) from Experiment 3 predicted by a mixed effects random intercept model grouped by rater with sex of rater as a second level predictor. (Observations: 3456, Groups: 72, AIC= 11370, BIC=11443)

	Value	SE	DF	t	p
Masculinity Intercept	2.0741	0.14935	3376	13.887	$< .0001$
Male Avatar	2.0625	0.12379	3376	16.662	$< .0001$
Male Confederate	0.3229	0.12379	3376	2.609	0.0091
Male Rater	0.1645	0.14405	70	1.142	0.2573
Has Audio	-0.1311	0.08753	3376	-1.497	0.1344
Has Motion	-0.1059	0.08753	3376	-1.210	0.2264
Male Avatar × Has Audio	-0.3455	0.10107	3376	-3.418	0.0006
Male Avatar × Has Motion	0.0799	0.10107	3376	0.790	0.4295
Male Confederate × Has Audio	0.8872	0.10107	3376	8.777	$< .0001$
Male Confederate × Has Motion	-0.0313	0.10107	3376	-0.309	0.7572

To better understand the interactions in these data, Figure 7 presents sex judgments and ratings of masculinity and femininity as interaction graphs broken down by male avatars and female avatars. Here one can see that the effects for video+audio and audio+still image are more similar to each other than to the effect for silent video. While motion has an effect by itself, it appears that once audio is present, there is no additional effect of motion.

Discussion

The viewers chose the sex of the person in the video clip to be the same as the sex of the avatar 91.9% of the time and the largest effect in the logistic model was that of the sex of the avatar, thus raters based their judgements of sex almost exclusively on the sex of the avatar. Thus, we consider the manipulation of sex to be an effective manipulation: The perceived sex of a conversant is almost always the sex of the avatar we display. When the sex of the avatar was not the judged sex, the mismatches were more likely to be female (4.6% female versus 2.8% male) which might be an indication that our avatars may be slightly feminizing, perhaps due to the appearance smoothing that is inherent in the AAM process.

Ratings of femininity were higher when the avatar was female, and when the viewed clip included motion or did not include audio. Thus, there appear to be independent contributions of both appearance and movement to the perception of femininity. Adding audio to the display tended to decrease the effect of the sex of the avatar and increase the effect of the sex of the confederate. One way to interpret this interaction is that the audio dynamics carried information about femininity that was not carried in the motion dynamics. This is particularly interesting since mouth movements are highly correlated with vocal dynamics.

Ratings of masculinity were higher when the avatar was male and when the confed-

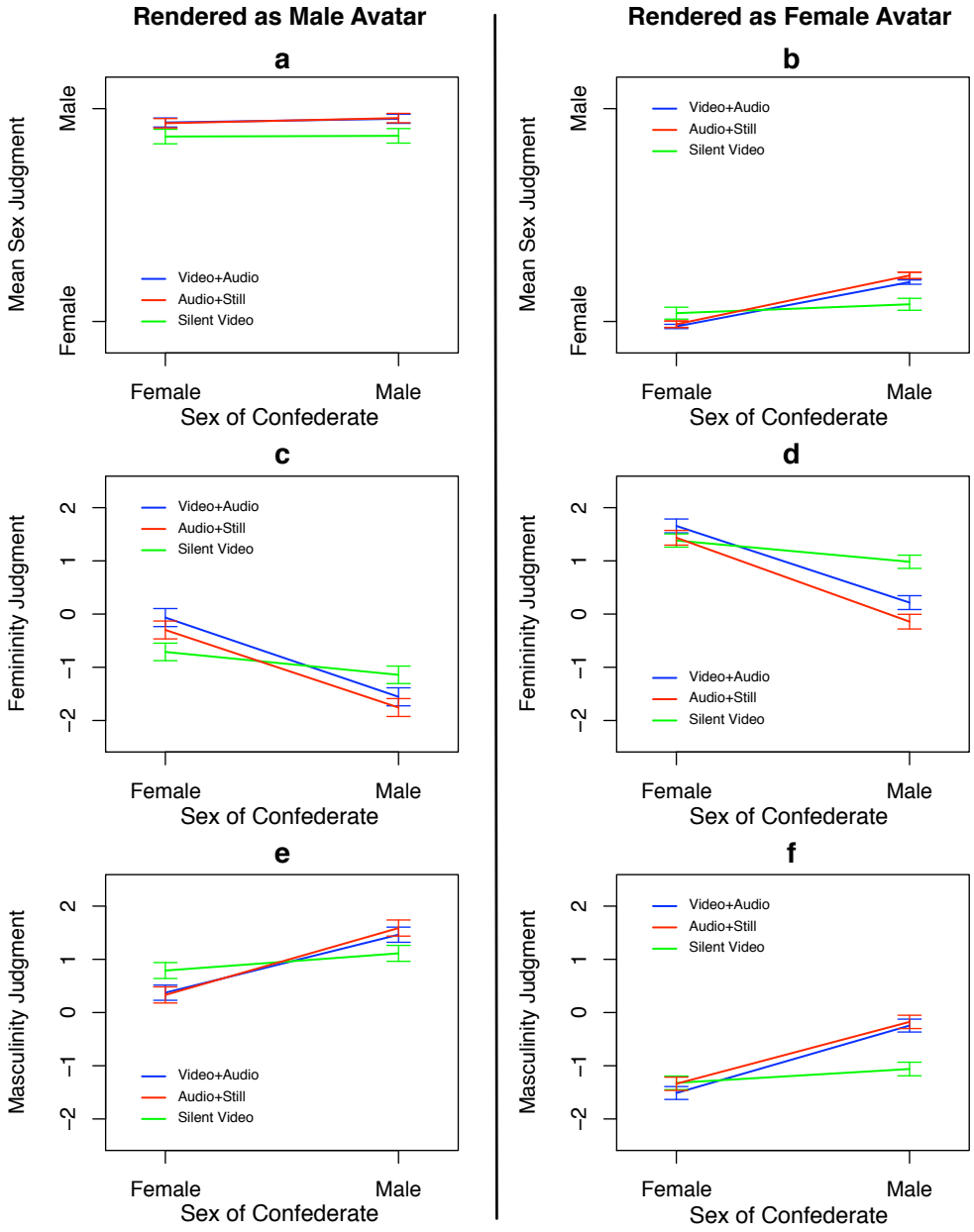

Figure 7. Judgments of Sex, Femininity, and Masculinity from 72 raters in Experiment 3. (a,c,e) The left column of graphs plot the Sex, Femininity and Masculinity judgments versus the sex of the confederate for video clips rendered as male avatars. (b,d,f) The right column of graphs plots the same judgment variables for the same video clips, but this time judges viewed the clips rendered as female avatars. Note: Error bars are 1.96 × SE.

erate was male. There were also interactions with audio that weakened the effect of the sex of the avatar and strengthened the effect of the sex of the confederate. Thus there were independent contributions of facial appearance and vocal dynamics to ratings of masculinity, but there did not appear to be an effect of motion dynamics. Again, this suggests that there were additional masculinity cues in the vocal dynamics that were not present in the facial dynamics.

We conclude from this experiment that judgments of sex are driven almost entirely by appearance while ratings of masculinity and femininity rely on a combination of both appearance and dynamics. For femininity judgments it appears that this is a combination of motion and vocal dynamic cues whereas for masculinity judgments it appears that the dynamics of the voice are primary.

General Discussion

As avatars approach being human–like in terms of appearance and movement, they risk being perceived as being more unnatural; what Mori termed the *uncanny valley* (Mori, 1970). Seyama and Nagayama (2007) investigated perception of static facial images and found that while this uncanny valley appears to exist, it is triggered by abnormal features. Wallraven and colleagues (1997) report that animation using motion capture data may compensate for reduced fidelity in form when participants rate variables such as sincerity and intensity of computer generated avatars. The resynthesized avatars we developed were accepted by naive participants as being video of another person viewed over videoconference and thus appear to have crossed the uncanny valley.

Using resynthesized avatars in a videoconference setting, we found that the amplitude and velocity of conversants' head movements is influenced by the dynamics (head and facial movement and/or vocal cadence) but not the perceived sex of the conversational partner. This finding is robust in that the apparent identity and sex of a confederate was randomly assigned and the confederate was blind to the identity and sex which they appeared to have in any particular conversation. Naive participants spoke with each confederate 3 times, so we were able to make strong conclusions about apparent sex given that the actual dyadic composition did not change while we manipulated the apparent dyadic composition. The manipulation was believable in that, when given an opportunity to guess the manipulation at the end of experiment, none of the naive participants was able to do so. Even when, in Experiment 1, naive participants were shown the avatar and full video in alternation, none of the participants guessed that the avatar was not what we said in our cover story, "video cut out around the face". The sex manipulation was effective in that in a follow–up experiment, 91.9% of the time raters chose the sex of the avatar as the perceived sex of an individual shown in a rendered video clip.

Thus, we conclude that gender–based social expectations are unlikely to be the source of reported gender differences in head nodding behavior during dyadic conversation. Although men and women adapt to each other's head movement amplitudes it appears that this may simply be a case of people (independent of sex) adapting to each other's head movement amplitude. It appears that a shared equilibrium is formed when two people converse. There were substantial individual differences in the level of that equilibrium as evidenced by the standard deviation of the random intercept for each of the measures of head movement. Given a shared equilibrium in a dyad, both Experiments 1 and 2 found

negative coupling between participants such that when one person moves more, the other person moves less. This result is a correlational result and not a manipulated result. However, the methodology we present would allow an experiment in which head amplitude could be manipulated in order to make a strong test of this hypothesis of negative coupling.

Overall, our results are consistent with a hypothesis of separate perceptual streams for appearance and biological motion (Giese & Poggio, 2003; Haxby et al., 2000). We find that head movements generated during conversation respond to dynamics but not appearance. We find that judgements of sex are influenced by appearance but not dynamics. Judgements of masculinity and femininity are more complicated, having independent contributions of appearance and dynamics. This dissociation of the effects of appearance and dynamics is difficult to explain without independent streams for appearance and biological motion.

Software Limitations and Future Directions

The software, while remarkably effective, has many areas that could be improved. In order to track and resynthesize a face, a model must be hand constructed from approximately 30 to 50 frames of previously recorded video. This process of model construction takes between 2 and 3 hours per participant. Thus, at the present time it is not possible to bring a participant into the lab and immediately track their facial movements — A preliminary session must be scheduled for video capture. In the current experiments we used confederates for the avatars, so this did not present a problem. However, if the experimental design involved tracking and synthesizing two naive participants, each dyad would require two sessions separated by a day and 4 to 6 hours of coding time to prepare for the second session. An improved generic model that tracked most people could allow a wider variety of experimental designs.

The current software is sensitive to contrast and color balance of the lighting in the video booth. The lighting during the original video recording, on which the model is built, must be the same as that when the participant is being tracked. We used studio lighting with stable color balance and a controlled lighting setting — the video booth in order to normalize the lighting. Improved software would automatically adjust for color balance and contrast in order to maximize the fit of the tracking part of the algorithm.

The identity transformations we applied in the current experiment are all between individuals of the same age. Older individuals have facial features such as wrinkles that move dynamically and do not appear on younger individuals faces. Wrinkles tend to vary across individuals and the folds that they make on two different individuals are not necessarily the same given the same contraction of facial muscles. It remains an open question whether the current model can be used to artificially age an individual or to switch identities between a younger person and an older person.

The AAM models used here do not track eye movements. Eye movements are, of course, a critical component of coordinated action during conversation. While our resynthesized models do correctly reproduce the movements of the eyes, they do so in the appearance model part of the AAM, and so we do not have information as to gaze direction. Ideally, we would like to be able to track eye movements as a source of data as well as manipulate eye gaze in the avatar. The resolution of the video camera is a limiting factor in tracking the center of the pupil. Higher resolution cameras and models that track eye movements are likely improvements to the methodology.

While we track the nonrigid movements of the face, it is an open question as to how these nonrigid movements are represented in perception. Using as few as 8 principal components of shape we have produced avatars that are convincing. This is a demonstration proof that the number of degrees of freedom in perceived in the dynamics of facial expression is likely to be much smaller than the number of independently controllable facial muscles. The current methodology does not guarantee that each principal component maintains the same meaning across individuals. This limits the usefulness of the AAM representation as a means of mapping facial expressions to affective dimensions. However, it may be that a confirmatory factor analysis approach may be able to applied to the data from the current experiments in order to better understand how the dynamics of facial expressions are related to perceptions of affect.

Finally, while the avatars we use move in a convincing manner, they are actually static models. That is to say, each frame of video is fit independently of previous video frames. Thus, a limitation of the experiments presented here is that they only manipulate identity, not dynamics. By manipulating dynamics, we could dissociate components of movement and vocalization that contribute to conversational coordination and person perception. A better avatar model would include both shape and appearance as well as a model for dynamics. The data from the current experiments will be used to further understand the parameters of the dynamics of biological rigid and nonrigid motion of the face. Using this understanding, we expect to be able to extend our model such that we can manipulate the dynamics of an avatar.

Experimental Limitations and Future Directions

Experiments 1 and 2 did not explicitly probe naive participants about how many individuals they spoke with or what the sex of those individuals might have been. Future studies would be strengthened by including such an explicit probe. The probe would need to come at the end of the session in order to not prime the naive participants to be looking for identity–based manipulations.

The overall means for RMS–H displacement and velocity in both Experiments 1 and 2 were substantially higher than observed in a previous face–to–face experiment (Ashenfelter et al., in press). We have yet to account for this phenomenon. This increase was observed in the unaltered video in Experiment 1 when the effect of the avatar display was taken into account. We expect this effect is likely due to some part of the experience of being in a videoconference, but we have not yet isolated what component or components of the videoconference setting lead to greater extent and velocity of head turns.

Rigid head movements are only a small part of the movements that are coordinated in conversation. Non–rigid facial expressions are certainly coordinated as well. The current analyses do not take into account coordination of facial expressions. Further analyses need to be performed to estimate the coordination between naive participants' and confederates' facial dynamics. Video from Experiments 1 and 2 could be tracked for non–rigid facial movement. However, analysis of these movements presents technical problems that are not currently solved.

The current analyses aggregate head movements over relatively long intervals of time. It would be informative to perform multivariate time series on both the rigid and non–rigid head movements from these experiments. In this way we may be able to better understand

the way that short term coupled dynamics lead to the longer term effects observed here.

An analysis of the vocal dynamics coupled with the analyses presented here may be able to shed light on how prosodic elements of speech influence rigid and non–rigid movement coupling between conversational partners. In addition, semantic analysis of the vocal stream may help shed light on how the non–verbal characteristics of the conversation help provide information in support of the semantic verbal stream.

Cultural, age, and ethnic differences between conversational partners may also play a role in the coordination of rigid and non–rigid head movements. Avatars that randomly assign these characteristics could be developed in order to test the extent of these effects. Finally, eye movements almost certainly play a substantial role in the dynamics of conversational interaction. Including eye tracking in future experiments would strengthen the measurement of the dimensions within which conversational partners organize their communication.

Conclusions

The avatar videoconference methodology can separate appearance and motion in natural conversation — Naive participants did not guess they were speaking with an avatar. Matched controls can be created such that confederates are blind to their apparent identity during a real time conversation, allowing random assignment of variables such as sex, race, and age that have been shown to carry implicit stereotypes. Judgment experiments can be performed where context of conversations is exactly matched across manipulations of identity, removing confounds due to context effects such as accent, vocal inflection, or nonverbal expressiveness of the speaker. We anticipate that this advance in methodology will enable new insights in dyadic and group interactions.

Simply stated, the main finding of these experiments is this: It is not what sex you appear to be, but rather how you move that determines how a shared dynamic of head movement is formed in a dyadic conversation. This result indicates that motion dynamics in everyday conversation, long acknowledged as important but rarely studied due to methodological difficulties, is long overdue for systematic inquiry.

References

Ashenfelter, K. T., Boker, S. M., Waddell, J. R., & Vitanov, N. (in press). Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. *Journal of Experimental Psychology: Human Perception and Performance*.

Atkinson, A., Tunstall, M., & Dittrich, W. (2007). Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*, *104*(1), 59–72.

Bernieri, F. J., Davis, J. M., Rosenthal, R., & Knee, C. R. (1994). Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin*, *20*(3), 303–311.

Berry, D. S. (1991). Child and adult sensitivity to gender information in patterns of facial motion. *Ecological Psychology*, *3*(4), 349–366.

Boker, S. M., Deboeck, P. R., Edler, C., & Keel, P. K. (in press). Generalized local linear approximation of derivatives from time series. In S.-M. C. . E. Ferrar (Ed.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue*. Boca Raton, FL: Taylor & Francis.

Boker, S. M., & Rotondo, J. L. (2002). Symmetry building and symmetry breaking in synchronized movement. In M. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 163–171). Amsterdam: John Benjamins.

Cappella, J. N. (1981). Mutual influence in expressive behavior: Adult–adult and infant–adult dyadic interaction. *Psychological Bulletin*, *89*(1), 101–132.

Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception–behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. Hassin, J. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 334–361). New York: Oxford University Press.

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception*, *34*(10), 1171–1180.

Condon, W. S. (1976). An analysis of behavioral organization. *Sign Language Studies*, *13*, 285–318.

Cootes, T. F., Edwards, G., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(6), 681–685.

Cootes, T. F., Wheeler, G. V., Walker, K. N., & Taylor, C. J. (2002). View-based active appearance models. *Image and Vision Computing*, *20*(9–10), 657–664.

Dixon, J. A., & Foster, D. H. (1998). Gender, social context, and backchannel responses. *Journal of Social Psychology*, *138*(1), 134–136.

Duncan, S. D., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*, 179–192.

Grahe, J. E., & Bernieri, F. J. (2006). The importance of nonverbal cues in judging rapport. *Journal of Noverbal Behavior*, *23*(4), 253–269.

Griffin, D., & Gonzalez, R. (2003). Models of dyadic social interaction. *Philosophical Transactions of the Royal Society of London, B*, *358*(1431), 573–581.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neuronal system for face perception. *Trends in Cognitive Sciences*, *6*(1), 223-233.

Helweg-Larsen, M., Cunningham, S. J., Carrico, A., & Pergram, A. M. (2004). To nod or not to nod: An observational study of nonverbal communication and status in female and male college students. *Psychology of Women Quarterly*, *28*(4), 358–361.

Hill, H. C. H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*(3), 880–885.

Hill, H. C. H., Troje, N. F., & Johnston, A. (2003). Range– and domain–specific exaggeration of facial speech. *Journal of Vision*, *5*, 793–807.

Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neuropsychological evidence. *Neuropsychologica*, *31*(2), 173–181.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology* (p. 451477). New York: Cambridge University Press.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, *99*(3), 422–431.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, *83*(1), 126–137.

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research*, *43*(18), 1921–1936.

Lafrance, M. (1985). Postural mirroring and intergroup relations. *Personality and Social Psychology Bulletin*, *11*(2), 207–217.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory and Cognition*, *27*(6), 974–985.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations model. *Journal of Personality and Social Psychology*, *65*(6), 1178–1187.

Macreae, C. N., & Martin, D. (2007). A boy primed Sue: Feature–based processing and person construal. *European Journal of Social Psychology*, *37*, 793–805.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: estimating the information employed for face classifications. *Cognitive Science*, *28*(2), 209–226.

Matthews, I., & Baker, S. (2004). Active appearance models revisited. *International Journal of Computer Vision*, *60*(2), 135–164.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, *7*(4), 33–35.

Morrison, E. R., Gralewski, L., Campbell, N., & Penton-Voak, I. S. (2007). Facial movement varies by sex and is related to attractiveness. *Evolution and Human Behavior*, *28*, 186–192.

Munhall, K. G., & Buchan, J. N. (2004). Something in the way she moves. *Trends in Cognitive Sciences*, *8*(2), 51–53.

Pollick, F. E., Hill, H., Calder, A., & Paterson, H. (2003). Recognising facial expression from spatially and temporally modified movements. *Perception*, *32*(7), 813–826.

Roger, D., & Nesshoever, W. (1987). Individual differences in dyadic conversational strategies: A further study. *British Journal of Social Psychology*, *26*(3), 247–255.

Rotondo, J. L., & Boker, S. M. (2002). Behavioral synchronization in human conversational interaction. In M. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 151–162). Amsterdam: John Benjamins.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! understanding recognition from the use of visual information. *Psychological Science*, *13*(2), 402–409.

Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence*, *16*(4), 337–351.

Steede, L. L., Tree, J. J., & Hole, G. J. (2007a). Dissociating mechanisms involved in accessing identity by dynamic and static cues. *Object Perception, Attention, and Memory (OPCAM) 2006 Conference Report, Visual Cognition*, *15*(1), 116–123.

Steede, L. L., Tree, J. J., & Hole, G. J. (2007b). I can't recognize your face but I can recognize its movement. *Cognitive Neuropsychology*, *24*(4), 451–466.

Wallraven, C., Breidt, M., Cunningham, D. W., & Bülthoff, H. H. (1997). Psychophysical evaluationof animatedfacial expressions. In *Proceedings of the 2nd symposium on applied perception in graphics and visualization.* New York: ACM Press.