Are you on my wavelength? Interpersonal coordination in naturalistic conversations

Jo Hale*, Jamie A Ward*, Francesco Buccheri, Dominic Oliver, Antonia F de C Hamilton

Institute of Cognitive Neuroscience, UCL, Alexandra House, 17 Queen Square, London, WC1N 3AZ

* Joint first authors

This is a preprint of a submitted paper. Released June 2018.

Short title

Interpersonal coordination in conversation

Author Affiliation:

Institute of Cognitive Neuroscience, UCL, Alexandra House, 17 Queen Square, London, WC1N 3AZ.

Orcid ID

Antonia Hamilton: http://orcid.org/0000-0001-8000-0219

Corresponding Author:

Antonia Hamilton, Institute of Cognitive Neuroscience, UCL, Alexandra House, 17 Queen Square, London, WC1N 3AZ, a.hamilton@ucl.ac.uk

Keywords:

Conversation, social coordination, motion capture, mimicry, synchronization, non-verbal

Abstract

Conversation between two people involves subtle non-verbal coordination but the parameters and timing of this coordination remain unclear, which limits our models of social coordination mechanisms. We implemented high-resolution motion capture of human head motion during structured conversations. Using pre-registered analyses, we quantify cross-participant wavelet coherence of head motion as a measure of non-verbal coordination, and report two novel results. First, head pitch (nodding) at 2.6 – 6.5 Hz shows below-chance coherence between people. This is driven by fast-nodding behaviour from the person listening, and is a newly defined nonverbal behaviour which may act as an important social signal. Second, head pitch movements at 0.2-1.1 Hz show above-chance coherence with a constant lag of around 600msec between a leader and follower. This is consistent with reactive (rather than predictive) models of mimicry behaviour. These results provide a step towards the quantification of real-world human behaviour and reveal mechanisms of social coordination.

Introduction

Face-to-face conversation between two people is the fundamental basis of our social interaction (Clark, 1996) and is of increasing interest in both clinical (Ramseyer & Tschacher, 2011) and neuroscientific (Schilbach et al., 2013) research. Despite this, we have only limited data on the precise timing and patterns of coordination which are found in real world conversation behaviours. Early work coding actions from videos shows that people tend to move coherently with one another, which has been described as mimicry or synchrony (Bernieri & Rosenthal, 1991). Behavioural studies suggest that this coordination acts as a ‘social glue’ (Lakin, Jefferis, Cheng, & Chartrand, 2003) that predicts the success of a negotiation or meeting (Pentland, 2008).

The initial aim of the present study was to use high precision motion capture to record the head motion of dyads in conversation, and then perform an advanced analyses to reveal the time course of mimicry. Our data show these features, but also revealed novel patterns of anti-coherence behaviour going beyond simple mimicry or synchrony. We conducted a pre-registered study to confirm the existence of this novel behaviour, and report here a detailed analysis of both mimicry in head motion and anti-coordination in head motion, as found in a naturalistic conversation task.

Background

It is increasingly recognised that neuroscience needs to examine real-world behaviour in detail (Krakauer, Ghazanfar, Gomez-Marin, Maciver, & Poeppel, 2017), and that social neuroscience needs to understand the interactions of two real people rather than just studying solo participants watching videos of people (Schilbach et al., 2013). A long tradition of research into human interpersonal coordination and conversation describes patterns of synchrony and mimicry as key behaviours in dyadic interactions. From early work (Condon & Ogston, 1966; Kendon, 1970) to the present (Chartrand & Bargh, 1999; Ramseyer & Tschacher, 2011), it is clear that people tend to move their heads and bodies in coordination both with each other and with their own speech rhythms. Untrained observers can perceive a gestalt of coordination in videos of conversation, and rate higher coordination in genuine interactions compared to ‘pseudo interactions’ where interaction partners from different videos were randomly paired together to make it look as though they were having a real interaction (Bernieri, 1988). Such studies clearly demonstrate that *something* important and interesting is happening in live dyadic interactions, but it is less clear exactly what this *something* is, and what mechanisms might support it.

A detailed understanding of dyadic coordination is important because it can constrain our theories of the mechanisms underlying such coordination. For a simple coordination between two people, we can imagine four different mechanisms with different timing properties. First, if each person is able to predict the other's action, then they will be able to show precise synchrony (0 msec lag) with both people making the same movement at the same time; this is seen in musical coordination (Konvalinka, Vuust, Roepstorff, & Frith, 2010) and implies a predictive mechanism. Second, mimicry could arise via an immediate reactive mechanism in response to another person, similar to visuomotor priming (Heyes, 2011); this type of mimicry might occur 300-1000 msec after a stimulus. Third, social mimicry might happen on much longer timescales of 2-10 seconds (Leander, Chartrand, & Bargh, 2012; Stel, van Dijk, & Olivier, 2009) which implies the involvement of a short-term memory mechanism. Finally, social behaviours might be governed by a particular pattern of phase relationships, such as in-phase movements being preferred (Oullier, de Guzman, Jantzen, Lagarde, & Kelso, 2008). If a constant phase-relationship is found across a range of motion frequencies, this implies a mechanism tuned to phase-lag rather than a fixed lag time. Thus, understanding the timing of social coordination and the frequency at which it occurs will help us understand the computational mechanisms which implement coordination.

Current data on the timing of mimicry and social coordination in real-world dyads are limited and of low resolution. It has been suggested that mimicry occurs with natural delays of 2-5 seconds (Leander et al., 2012; van Baaren, Decety, Dijksterhuis, van der Leij, & van Leeuwen, 2009), although estimates vary from 0 to 10 seconds (Chartrand & Bargh, 1999; Stel et al., 2009). This implies movement coherence within a frequency band of roughly 0.5 – 0.2 Hz. Coordination has also been linked to the natural rhythm of speech (Condon & Ogston, 1966; Hadar, Steiner, Grant, & Rose, 1983a, 1983b; Kendon, 1970), which is around 5 Hz (Morrill, Paukner, Ferrari, & Ghazanfar, 2012; Ohala, 1975). Further research suggests that facial expressions at frequencies of approximately 2 - 7 Hz are also important for the interpretation of speech (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). Other studies have demonstrated entrainment in covert rhythms such as heart rate (1 – 1.7 Hz; Konvalinka et al., 2011) and breathing (0.2 - 0.3 Hz; Pellegrini & Ciceri, 2012; Warner, 1996). Human body movement is also thought to be tuned to frequencies around 2 Hz based on musical preference and recall of these tempi (Noorden & Moelants, 1999). Based on these results, it is hard to make firm predictions about the frequencies at which people naturally coordinate, so in the present study, we examine coordination of head moment over the range from 0.2 to 8Hz, covering the frequencies of behaviours previously described.

Studies of human behaviour in conversation have traditionally been based on video of behaviour which can be coded by trained observers to understand the detail of what participants do. However, hand coding of videos in these studies may introduce biases (Cappella, 1981) and limits the amount of data which can be processed as well as its resolution (Grammer, Kruck, & Magnusson, 1998). More recently, researchers have begun to use automated methods to calculate motion energy or other parameters from video (Fujiwara & Daibo, 2016; Paxton & Dale, 2013; Ramseyer & Tschacher, 2010; Schmidt, Morr, Fitzpatrick, & Richardson, 2012), but resolution is still limited to the pixels which change in a flat image. Motion capture provides much higher resolution (Feese, Arnrich, Tröster, Meyer, & Jonas, 2011; Poppe, Van Der Zee, Heylen, & Taylor, 2013), so we use this method here. To extract meaning from the rich motion capture data, we use wavelet analysis in which each motion trace is represented in terms of wavelets at different frequencies and time points. The cross-wavelet coherence between the wavelets of two different people gives a measure of the time-frequency coordination between their movements (Fujiwara & Daibo, 2016; Issartel, Bardainne, Gaillot, & Marin, 2014). Proof of concept studies using wavelet methods have found that coordination exists during musical improvisation (Walton, Richardson, Langland-Hassan, & Chemero, 2015), and telling knock-knock jokes (Schmidt, Nie, Franco, & Richardson, 2014). Such studies suggest that coordination occurs at multiple timescales (Schmidt et al., 2014) and frequencies (Fujiwara & Daibo, 2016) with less coherence at high frequencies close to 4 Hz. However, previous studies have not always used detailed analyses nor performed the most robust comparisons between real dyads and pseudo dyads.

The present study

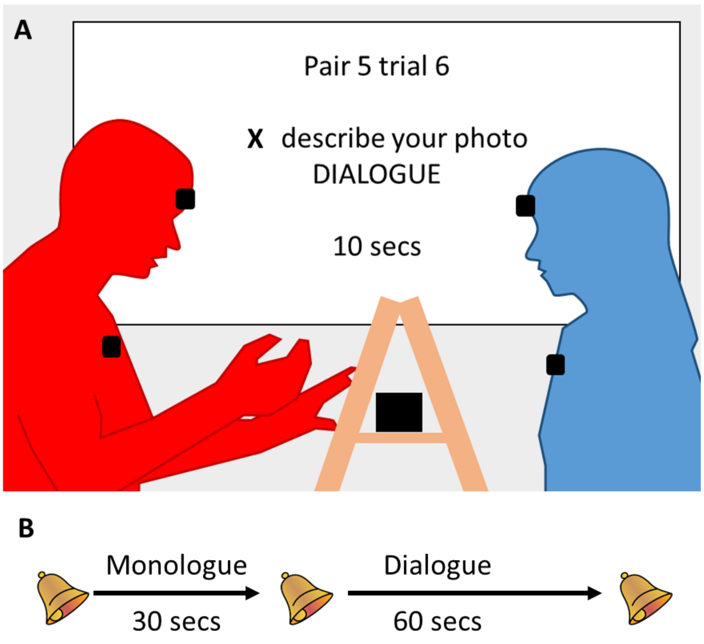

Figure 1. A. Data collection. Two participants set approx. 1m apart on small stools. A wooden frame between the participants held the picture (for description) and the Polhemus transmitter box. Polhemus markers were placed on the chest and head of each participant. A projector screen beside the participants provided instructions, timing cues and video synchronisation. **B. Trial structure.** Each trial had 30 seconds of monologue followed by 60 seconds of dialogue marked by 'ding' sounds. Participants completed 16 trials, alternative turns as leader / follower.

In the present study we recorded movements from pairs of participants (dyads) engaged in a picture description task (Figure 1). This task has been used previously in behavioural (Chartrand & Bargh, 1999) and mocap studies (Shockley, Santana, & Fowler, 2003). In each trial of our task, one participant had the role of Leader, holding a picture of a complex scene while the other had the role of Follower (Figure 1). Each trial lasted 90 seconds; for the first 30 seconds the Leader described the picture and the Follower remained silent (monologue section); for the remaining 60 seconds the Follower could ask questions and both participants could converse together about the picture (dialogue section). We recorded the position and rotation of motion sensors on each participant's head and torso, at a rate of 60 Hz. We use wavelet coherence measures to quantify the interpersonal coordination at different frequency bands from 0.2-8 Hz, for three distinct head motion signals: head yaw (turning), head pitch (nodding), and head roll (tilting). As wavelet analysis is a relatively new technique in social cognitive research, the present study consisted of a pilot phase and a final phase. The pilot phase was highly exploratory and the data was used to develop and test our analysis algorithms. We then pre-registered our methods and collected a second independent dataset (n=31 dyads), which we report here. Our aim was to precisely characterise the frequency, phases and patterns of head motion coordination in a conversation, and to test cognitive models of this coordination.

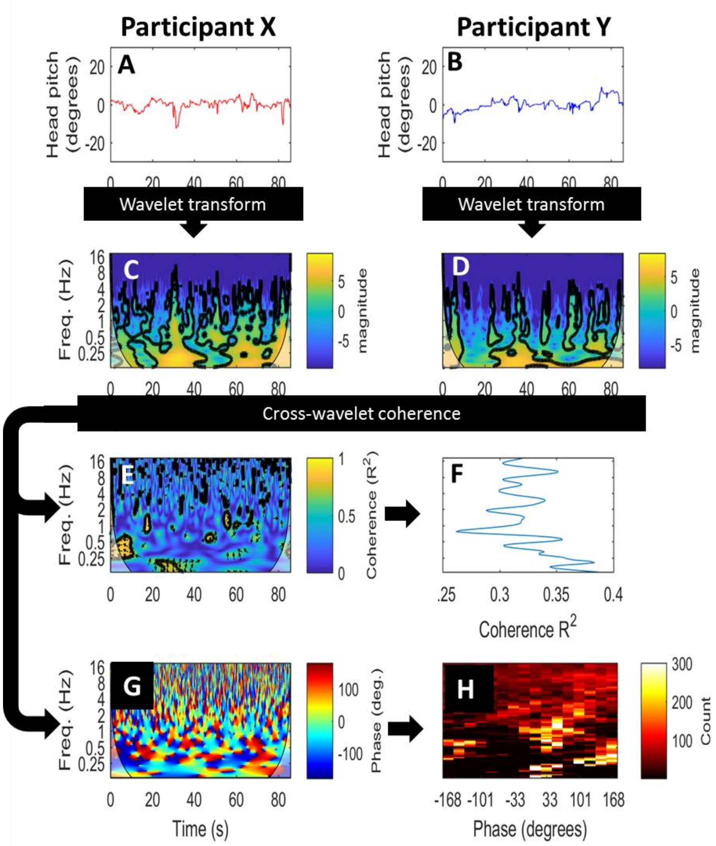

Figure 2. Data analysis methods. For each trial, the head motion traces for the two participants (A,B) are subject to a wavelet transformation giving C and D. The cross-wavelet coherence is calculated and quantifies coherence (E) and phase angle (H). The coherence is averaged over time (F) and then over all dyads to obtain the overall coherence measure. The phase angles are binned by angle and frequency (collapsing over time) (J) and then averaged over all dyads. For pseudo-coherence calculations, the same procedure was used with mis-matched data replacing the inputs A and B (see Supplementary Fig 2).

Results

Novel patterns of Cross-wavelet coherence of head movements in real vs. pseudo interactions (pre-registered analysis)

Each dyad completed 16 trials of the conversation task, 8 with person "X" as the leader, and 8 with "Y" as the leader. For each trial and each head motion parameter (yaw, pitch, roll), we calculated the wavelet transform and then the wavelet coherence (Figure 2). To provide a null contrast to the true interactions, we calculated the wavelet coherence for pseudo interactions where we shuffle the data within pairs and within leader/follower roles. For example, if trials 1, 3, 7 and 9 have X as leader and Y as follower, the true trials are 1-1; 3-3; etc, while the pseudo trials are 1-3; 7-3; etc. This strict definition of pseudo trials within-pair and within-role gives the strongest possible test that any differences between the coherence levels in the real and pseudo pairs must be due to the live interaction and not to differences in participants or roles.

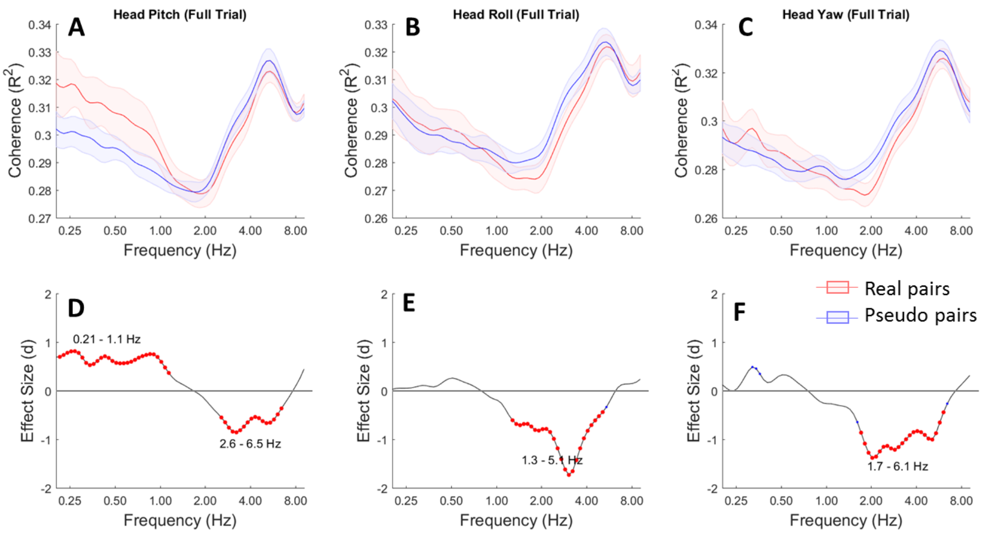

Figure 3. Wavelet coherence across different motion signals. A, B and C show mean and standard error of coherence across pitch, row and yaw for real pairs (red) and pseudo-pairs (blue) of trials. D, E and F show the effect size for a t-test on the difference between real and pseudo pairs, with red dots indicating points that pass a p<0.05 FDR significance threshold.

Levels of coherence at each motion frequency are shown in Figure 3. The top row of the figure shows the mean coherence in the real trials (red) and pseudo trials (blue). To assess the difference in cross-wavelet coherence between real and pseudo interactions, we performed t-tests at each frequency (89 tests). The plots in the lower row show the effect size (Cohen's *d*) for the difference between real and pseudo interactions, plotted for head pitch, roll, and yaw respectively. Red dots indicate frequencies where there was a significant difference between real data and pseudo data, with FDR correction for multiple comparisons (Benjamini & Hochberg, 1995). Two patterns are noticeable in this data, especially with head pitch (Figure 3A&D). First, there is greater coherence in real pairs than pseudo pairs at low frequencies (significant for head pitch at 0.21 – 1.1 Hz). We term this 'low frequency mimicry' and examine the phase & lag patterns in this frequency range below. Second, there is a marked dip in coherence in the real pairs at higher frequencies (significant for pitch at 2.6 – 6.5 Hz), compared to coherence between pseudo pairs. We term this 'high frequency anti-coherence'.

The high frequency anti-coherence pattern was entirely unexpected and has never (to our knowledge) been described before. When we first saw this pattern in our pilot data (Hale, 2017), we worried about its robustness and so conducted a replication study with a larger sample size and a pre-registered data analysis pathway (Hale & Hamilton, 2016). It is the pre-registered results which we report here (Fig 3). These data show that the high frequency anti-coherence is reliable and replicable. We have conducted further exploratory analyses to understand the origins of this unexpected behaviour.

Exploring high frequency anti-coherence in head motion.

To understand the anti-coherence pattern, it helps to clarify what this data actually means. In a cross-wavelet analysis, two signals have high coherence if both have energy at

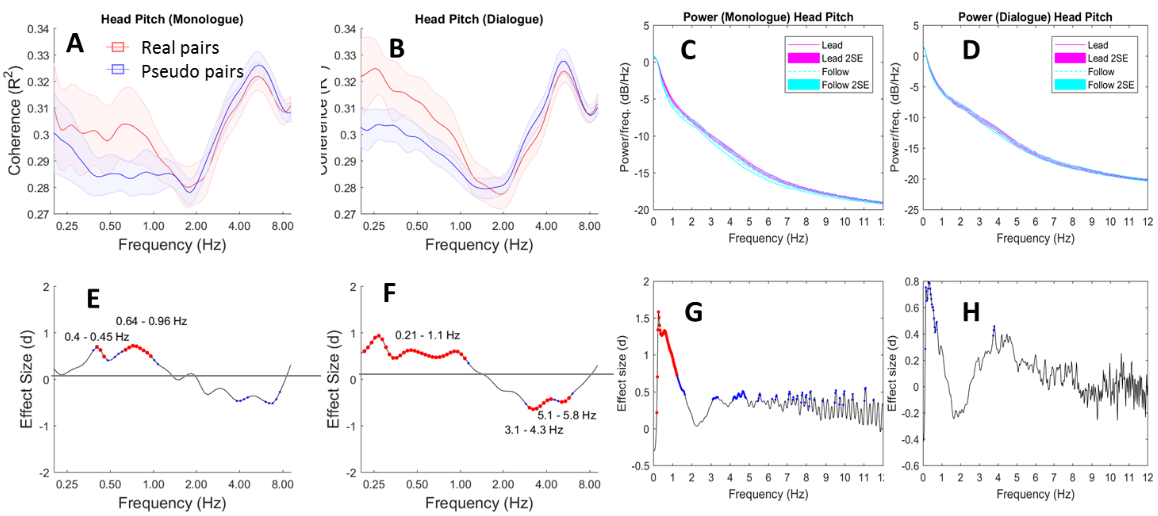

Figure 4. Global differences between monologue-dialogue phases and between leader / follower power do not drive the anti-coherence effect. A and B illustrate the wavelet coherence for monologue and dialogue segments of the pitch data, and E and F illustrate the effect size for the same data (with the same conventions as Fig 3). C and D illustrates the power at each frequency in the leader and follower data (for monologue and dialogue respectively), and G and H illustrates the effect size for a difference in power between the two roles (for monologue and dialogue). In all cases, red dots indicate effects which are significant at p<0.05 FDR corrected.

the same frequency (Issartel et al., 2014), regardless of the phase relationship between the two signals. Anti-coherence in a wavelet analysis means that two people are moving at different frequencies, for example, when one moves at 2.5Hz, the other moves at 3.5Hz. It is not the same as anti-phase coherence, where two people move at the same frequency but out of phase which each other. We present a number of exploratory analyses to understand the characteristics of the anti-coherence occurring at frequencies broadly between 2.6 and 6.5 Hz. The results we report here focus on the head pitch data, where the anti-coherence effect was strongest. Equivalent analyses on roll and yaw are presented in supplementary information.

First, we tested if this anti-coherence was present in both the monologue and dialogue phases of the conversation. For example, if anti-coherence arose because the Leader's head moved at 3Hz while speaking but the Follower remained still, then we might expect the signal to be clearer in the monologue phase of the trial compared to the dialogue. Figure 4A&E illustrates that the anti-coherence effect does not reach significance when examined in the monologue phase only, but is reliable in the dialogue phase (Fig 4B&F). This suggests that anti-coherence is a general property of conversation and is not limited to the more artificial context where one participant is not allowed to speak.

Second, we checked if the anti-coherence could be driven by global changes in the signal power at each frequency. For example, if one participant moved much less than the other at 3Hz, then there would be no power to drive coherence at 3Hz. As we have used a strict, within-dyad comparison between real trials and pseudo-trials, it is unlikely that power differences drive our effects, because data with similar power occur in both the real and pseudo trial analyses. To double-check this, we calculated spectral power over the whole trial (and compared power spectrum between leader and follower roles. In the monologue case (Fig 4C&G), we show that leaders have more power than followers at low frequencies (the range where we identified low-frequency mimicry). This corresponds to the fact that leaders do all the talking during monologue. But at higher frequencies, and throughout most of the dialogue case, as shown in Fig 4D&H, the leader and follower power levels are evenly matched.

These results suggest that the higher-frequency anti-coherence seen in head motion between two people in live conversation is not driven by global differences in signal power or by changes to the task context from monologue to dialogue. Rather, there must be a more subtle change in behaviour. To explore this, we identified segments of conversation where one participant is moving at a frequency between 1.5 and 8Hz (ie. above the cross-over point in Fig 3D) and label these as 'fast nods'. We then characterise them in relation to what the other participant is doing, as outlined below.

Detecting and understanding fast nods

To detect fast-nods in the head motion data of a single participant, we built a simple detector which estimated the dominant frequency in head pitch using a zero-crossing method (see method section). Using this detector, we coded each 1-second window in our data as containing a fast nod from the leader, the follower, both, or neither. Figure 5 A illustrates the fast nods detected in dialogue from one sample trial, classified by Leader and Follower roles. We find that followers spend 22% of the trial engaged in fast-nod behaviour while leaders spend only 10% of the trial engaged in this behaviour, and this difference is significant (t(25)=5.83, p<0.00001, effect size d = 1.394). Figure 5 B illustrates this result. In terms of classifier precision (where precision = nodding time as follower / total nodding time), 67% of any reported fast nods are performed by a person who is in the follower role.

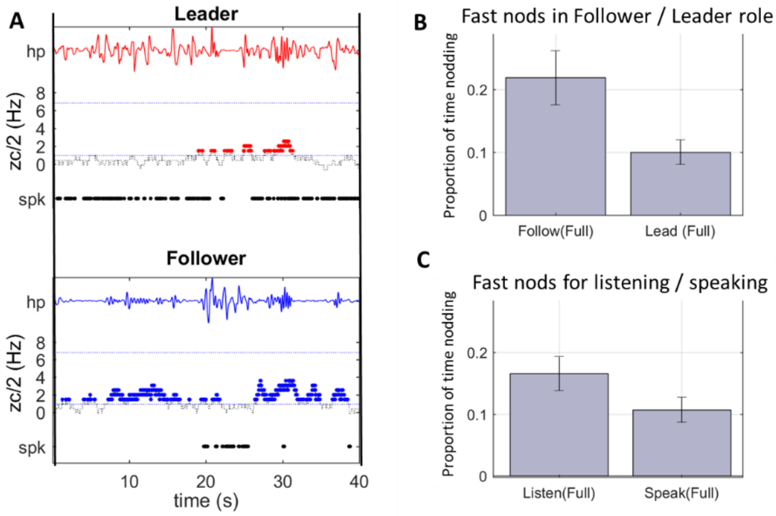

Figure 5. Detection of fast nods & their relationship to speech. A is a sample timeline from a leader & follower in the same trial, illustrating the head pitch signal (hp), the output of the 'fast nod detection algorithm' (zc/2) and whether that person is speaking or not (spk). Fast nods (coloured dots) are more common when a participant is listening. This is seen when comparing fast-nods categorised by the role of Leader v Follower over a whole trial (B) and also when comparing the rate of fast-nods when an individual is speaking or not-speaking (determined for each time point within a trial) in panel C.

This effect suggests that fast nodding is probably related to listening behaviour, since the role of the Follower is to listen to the Leader. However, it is possible to examine the relationship between listening and fast-nodding in more detail by using audio recordings of each trial. We recorded audio signals from each participant onto the left and right channels of a single audio file, and thus can classify who was speaking at each time point in the trial using a simple thresholding of signal energy in the left and right channels. We then test, for each trial, how the speaking behaviour of X (speaking / not speaking) predicts the fast-nodding behaviour of Y (nodding / not nodding) and vice versa, irrespective of leader/follower roles. We find that participants are more likely to be nodding when their partner is speaking than not (t(23)=4.08, p<0.001, effect d size = 0.843). Figure 5 C plots this difference in terms of the proportion of fast nods Y made while their partner X was speaking (17%) or not speaking (11%). The precision of this classifier is a high 94%. This

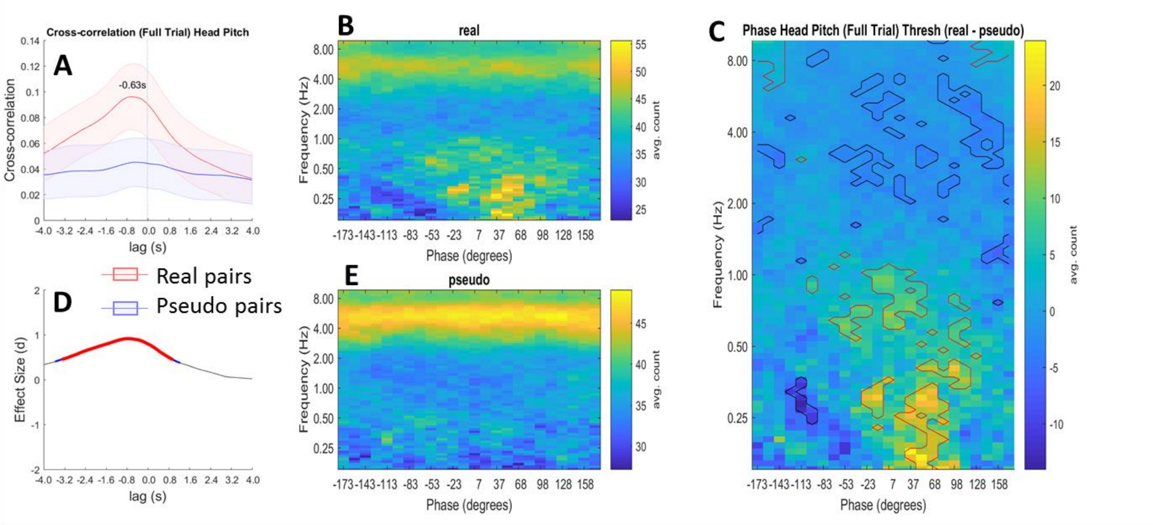

Figure 6. Phase angle analysis. A. A cross-correlation analysis of head pitch data reveals a peak at -0.6 seconds, consistent with the speaker's head pitch predicting the listeners head pitch by this amount. The effect size and t-values are illustrated in D. Phase-frequency plots for real (B) and pseudo (E) data provide a more detailed illustration of this phenemona. In the difference plot (C), regions with positive effect (more data in the phase bins for real pairs) are outlined in red and regions with a negative effect (more data in phase bins for pseudo pairs) are outlined in blue.

means that if one person is fast-nodding, it is extremely likely that their conversation partner is speaking.

Exploring the characteristics of low frequency mimicry

In addition to the fast nods described above, our data also revealed a positive coherence between dyads at 0.2 – 1.1 Hz in head pitch. These frequencies are commonly linked to mimicry behaviour, and this positive coherence is a plausible signature of the mimicry of one participant's head movement by the other. As before, we focus primarily on head pitch. In this data, head yaw and roll did not reach FDR significance for positive coherence at any frequency. To understand the mechanisms of this behaviour, we first tested how the positive coherence related to conversation in terms of monologue / dialogue segments and who was speaking at each time. Figure 4A&E illustrates that the coherence effect is partly significant (between 0.4 – 0.96 Hz) when examined in the monologue data, and is consistently significant (between 0.2 – 1.1 Hz) in the dialogue data (Fig 4B&F). This means that low-frequency coherent head motion is seen in both monologue and dialogue segments of the conversation.

A key question in examining this mimicry behaviour concerns the time delays or phase lags between participants – do they move synchronously, with a fixed time delay or with a fixed phase delay? Our first analysis of this question examined the cross-correlation between the Leader's head pitch and the follower's head pitch over each whole trial. Leader to Follower cross-correlation was calculated across a range of different lag times (-4 to 4 s) for real and pseudo trials. These results are averaged and shown with standard error (Figure 6A), and with Cohen's-d effect size (Figure 6D) for the comparison between the real and pseudo conditions. We find that real trials have greater cross-correlation than pseudo trials across a range of lags from -3 to 0.8 seconds, with a peak at -0.6 seconds. This implies that the follower tends to match the head movements of the leader with around a 600 msec delay.

We also examined the phase relationship between the two participants for the regions of the frequency spectrum with positive coherence. That is, for each wavelet coherence plot of each trial, we calculated the phase relationship between leader and follower at every time-frequency point (see methods). We plot these as histograms of phase-angle counts (quantized into 24-bins) for each frequency band in Figure 2 G&H with movement frequency on the y axis and phase-bins on the x axis in Fig 2H We averaged the phase histograms over trials and participants for both real trial pairs and pseudo trial pairs. Figure 6 shows the combined phase plot over all trials for real (Fig. 6 B) vs. pseudo pairs (Fig. 6 E) with the difference between these in Fig 6C. The highlighted areas of positive significance ($p<0.05$ in a paired t-test, df=25) show that coherence occurs in the 0.2 – 0.5 Hz frequency range with approximately 30-90 degree phase shift between leader and follower. This means that the head motion of the participant holding the picture leads the head motion of the follower by 30-90 degrees of phase. Note also that the dark blue areas in the top half of Figure 6C reflect the high-frequency anti-coherence described above, confirming that this pattern can also be seen in a different analysis.

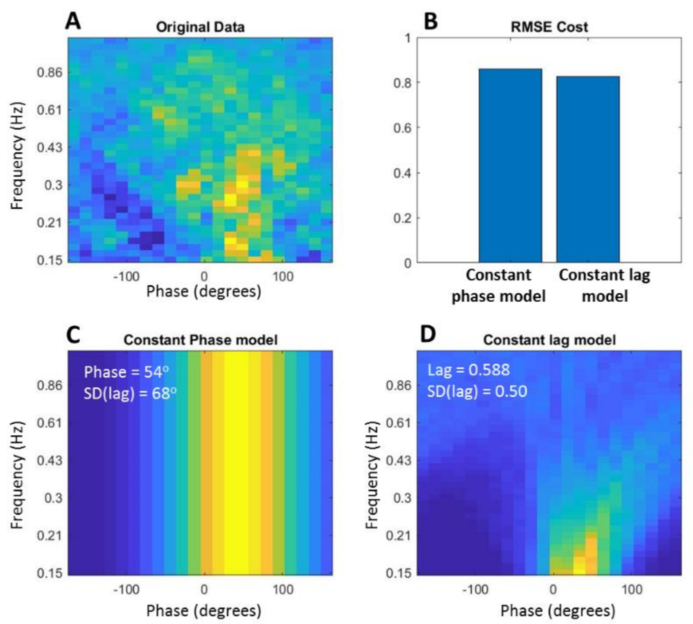

Figure 7. Modelling the phase-frequency relationship. The original data in phase-frequency histogram space (A) was modelled with a constant phase model (C) and a constant lag model (D). Optimal model outputs & their parameters are shown in C and D. The latter gave a better fit to the data (B).

Two different mechanisms could potentially drive the phase effect shown in Fig 6C. Followers could be in sync with leaders, maintaining a specific phase relationship (e.g. 40° of phase) across a range of frequencies ('constant-phase' mechanism). Or followers could lag behind leaders with a fixed delay (e.g. 600 ms; 'constant-lag' mechanism). The appearance of Fig 6C, with slightly greater phase shifts for slightly higher frequencies, suggests the latter explanation. To test this formally, we built simple models of the two potential mechanisms – a constant phase model and a constant lag model (see methods). Each model had two free parameters and which were optimised with Matlab's fminsearch. The results with the optimal parameters are shown in Figure 7. The RMSE of the constant-lag model was lower than the constant-phase model, indicating that this gives a better explanation of the data. The optimal constant-lag parameter of 0.588s is close to the mean lag of 0.63s in the covariance analysis of Figure 6 A. Thus, both analyses support the idea that leader-follower head motion can be characterised by a constant-lag model with a lag of approximately 0.6 seconds.

Discussion

This paper presents a detailed analysis of a rich dataset describing head motion of dyads during a structured conversation. Using high-resolution recordings and robust advanced analysis methods, we find evidence for important motion features which have not previously been described. First, real pairs of participants show head pitch (nodding) coherence at low frequencies (0.2-1.1Hz) and this coherent motion has a timing lag of 0.6 s between leader and follower. Second, significant anti-coherence in head pitch is seen at high frequencies (2.6-6.5 Hz frequency band), with evidence that this is driven by fast-nodding behaviour on the part of listeners (not speakers). We consider what this means for our models of human social interaction and for future studies.

Low frequency coherence

Our data reveals low-frequency coherence between 0.2 and 1.1 Hz with a time lag around 600ms between speaker and listener. This pattern is consistent with many early reports of mimicry (Condon & Ogston, 1966) as well as more recent motion capture studies (Fujiwara & Daibo, 2016; Schmidt et al., 2014). The frequency range below 1.1 Hz is also consistent with timescales traditionally associated with behavioural mimicry (Chartrand & Bargh, 1999; Stel et al., 2009). The lag of around 600 ms suggests that this behaviour is not driven by an anticipatory mechanism which perfectly synchronises the motion between two people. Such mechanisms can be seen in tasks involving musical rhythms (Konvalinka et al., 2010) but do not seem to apply here. We also do not see clear evidence for coherence at long time scales (2-5sec lag) which might imply a role for memory in producing mimicry, though we note that our trial duration of 90 seconds might limit our ability to detect coherence over timescales as long as 10 seconds. Rather, the low-frequency coherence we find has a time-lag around 600 msecs, and fits better to a constant-lag model than to a constant-phase model. This pattern of behaviour is consistent with a reactive mechanism in which one person sees an action and does a similar action around 600 msec later, which can be implemented by simple visuomotor priming (Heyes, 2011). There is no need for either prediction or memory of the other's action in explaining the patterns of behaviour we see.

Head pitch data showed the strongest pattern of coherence among the signals we measured (pitch, yaw and roll), and this pattern was seen in both monologue and dialogue. In terms of simple behaviours, one possible description is that followers mimic the head posture and nods of the other (Chartrand & Bargh, 1999). A related possibility is that leaders in this picture description task must alternate their gaze between looking at the face of their partner and looking down at the picture which is being described, and followers may copy this gaze pattern (Friesen & Kingstone, 1998; Tomasello, 1995) even though they cannot see the picture. Our data does not discriminate between 'pure mimicry' and 'gaze following' but both are interactive social behaviours which occur in true interactive dyads and not in pseudo data or solo behaviour. In our data, both are also compatible with a simple reactive model rather than a prediction model or memory model. The high resolution data recordings in the present data set allow us to explore the characteristics of real-world interactive behaviour and to precisely parameterise it.

High frequency Anti-coherence

The most surprising finding in our data was a pattern of high frequency (2.6 Hz – 6.5Hz) anti-coherence, whereby the two participants show systematically less-than-chance coherence in their head motion. Note that anti-coherence is not the same as coherence in anti-phase, where two people move at the same frequency but out of phase; here, participants do not move at the same frequency at all. Previous data suggests that head movements in the frequency range of 2.6 – 6.5 Hz are fairly typical in conversations. Hadar et al. (1983b) recorded head rotations while dyads engaged in free conversations and found that 'ordinary' head motion occurred within a range of 1.8 – 3.7 Hz, while 'fast' movements were characterised as above 3.7 Hz. Although their study was based on a very small sample size (N = 4), Hadar et al.'s account suggests that the decoupling we observed was within a normal to fast range for head movements in conversation. Therefore, given the wealth of evidence that people spontaneously coordinate other movements (Bernieri, Davis, Rosenthal, & Knee, 1994; Grammer et al., 1998; Ramseyer & Tschacher, 2010; Schmidt et al., 2012), it is surprising we should see active decoupling of head movements at typical frequencies for conversation. In fact, we were so surprised by this result in our pilot data that we ran the current replication with a fully pre-registered analysis pathway to verify the existence of the anti-coherence, and found that it is a robust effect. Further exploratory analysis shows that anti-coherence is tightly linked to each person's speaking or listening status. We found that leaders (who mainly speak) and followers (who mainly listen) nod with similar energy between 2.6 - 6.5 Hz when data is averaged over whole trials (Fig4 C and D). However, there also a specific fast-nod pattern at 2.6-6.5Hz which occurs in the person who is listening (Fig 5). A simple zero-crossing estimation of dominant frequency in this range allowed us to classify a participant as a follower with 67% precision, and as listening to their partner speaking (based on audio recordings) with 94% precision.

We term this listening behaviour a 'fast nod' and consider several possible behavioural explanations of it. 2-5Hz is a frequency band dominated by speech rhythms (Chandrasekaran et al., 2009; Hadar et al., 1983b), so if fast-nodding at this frequency occurred in the speaker (rather than the listener) then we would consider it a side effect of jaw & vocal movements during speech. The fact that it occurs in the listener is thus rather surprising. One possibility is that the listener is attuning to the speech rhythms of the speaker and reflects this in fast nodding, possibly to aid comprehension. For example, seeing natural head motion aids speech comprehension (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004) and it is possible that performing movements also contributes. This could be tested in future studies by exploring the relationship between speech rhythm and head motion. A second possibility is that fast-nodding reflects a communication back-channel (Clark, 1996), whereby the listener is sending a signal of attentiveness and engagement to the speaker. Many previous studies also report a 'nodding' backchannel in human communication (Kendon, 1970). However, our result is novel because this is not a slow nod that might be recorded by a video camera, but is a very small and quick head motion, occurring at roughly the same frequency as human speech. To test this, further studies would need to determine if listeners produced more fast-nodding when with another person (compared to alone) and if speakers can detect the fast-nodding behaviour (on some level).

One possible criticism of our fast-nodding result is that it is not easy, voluntarily, to produce a small head nod at 2-5 Hz – certainly not at the 5 Hz end of this frequency band. We agree that this behaviour cannot easily be produced on command, but that is true of many other socially meaningful signals including genuine (Duchenne) smiles (Ekman, Davidson, & Friesen, 1990; Gunnery, Hall, & Ruben, 2013) and genuine laughter (Lavan, Scott, & McGettigan, 2016). The lack of a voluntary pathway for a particular behaviour does not mean that this behaviour cannot have an important role in social signalling. Rather, it makes that behaviour hard to fake and means that it may have more values as an honest signal of listening / attentiveness (Pentland, 2008). Further studies will be required to test this.

Limitations & future directions

Our data provides a novel, high resolution recording of pairs of naïve participants performing a structured conversation task. We specifically used large sample sizes and a pre-registered analysis pathway to ensure the robustness of our novel findings. Nevertheless, there remain some limitations to our data, and our project opens up many future directions for further studies. First, we explored only one task (picture description) which may induce artificial up-down head movements to look at the picture, and it remains to be seen if all the mimicry and anti-coherence patterns which we report here generalise to other tasks. Second, the audio data we recorded was of rather low quality – data was sufficient to detect who was speaking but not to allow a detailed analysis of speech rhythms, so further studies will be required to determine if head motion anticoherence is linked to specific speech rhythms. Third, our trials were only 90 seconds long, with only 60 seconds of dialogue which limited our ability to detect mimicry behaviour over long time windows (around 10 seconds). Finally, we recorded only head motion, and a richer dataset involving capture of hand motions, gaze and facial expressions would allow an even more detailed characterisation of human conversation behaviour. Setting the standard for such datasets and defining the way to analyse this data will allow us to explore individual differences in conversation behaviour and how clinical populations differ from typical populations, creating a new science of ecologically valid social interactions.

Conclusions

This paper describes a rich dataset of head movements in naturalistic conversations and provides new insights into the patterns and mechanisms of social coordination. We show a reliable anti-coherence at high frequencies (2.6 - 6.5Hz) which may be linked to fast-nodding behaviour on the part of the listener. We also show coherent head motion at low frequencies (0.21 – 1.1 Hz) with a time lag around 600 msec. This implies a simple reactive visuomotor mechanism which is likely to reflect a combination of mimicry and gaze following. The present study builds on a growing literature exploring interpersonal coordination through automatic motion capture and spectrum analysis. Such detailed analysis of the 'big data' of human social interactions will be critical in creating a new understanding of our everyday social behaviour and the neurocognitive mechanisms which support it.

Materials and methods,

Participants

For this study, we recruited 31 dyads and after data exclusions (see below), we report results from 26 dyads (M_{age} = 22.3 years, SD_{age} = 2.9 years). There were 16 same-gender dyads and 10 mixed-gender dyads (34 female and 18 male participants). Participants were recruited from a local mailing list and paired up as dyads on the basis of their availability and preference for the same time slot. All participants gave written consent and were paid £7.50 for 1 hour. The study received ethical approval from the UCL Research Ethics Committee and was performed in accordance with the 1964 Declaration of Helsinki.

Lab setup

The room was set up with two wooden stools for the participants, facing each other at a distance of approximately 1.5m. Between the stools, a wooden frame held a Polhemus transmitter device which generated an electromagnetic field of approximately 2m diameter around the participants. A projector screen to the side of the participants showed instructions throughout the session, and audio speakers behind the projector provided audio cues. A curtain separated participants from the experimenter, who remained in the room but did not interact with participants during the experiment and could not be seen.

A Polhemus magnetic motion tracking device (Polhemus Inc., Vermont) recorded the head and torso movements of each participant at 60Hz. One sensor was fixed on the participant's forehead using a Velcro cloth band. Another sensor was attached to the participant's upper back using microporous tape. Both participants wore a lapel microphone and their voices were recorded on the left and right channels of a single audio file. A video camera mounted on a tripod recorded the session, offering a clear view of the participants' seated bodies. We used Vizard software (WorldViz, California) to display instructions on the projector screen, trigger audio cues and record data.

Procedure

Pairs of participants arrived and gave informed consent to take part. They were fitted with motion sensors and microphones, and randomly assigned to be the 'X' participant or 'Y' participant. These labels were used to distinguish dyad members during the experiment and in the recorded data. Participants completed one practice block followed by four experimental blocks. Each block was made up of a short language task, followed by four trials of the picture description task. The language task was a social priming task as used in Wang & Hamilton (2013), which was designed to activate prosocial or antisocial concepts in the participants. Both participants completed the same type of priming (prosocial or antisocial) with different exemplars. As the priming had no discernible impact on behaviour in the task, we do not discuss it further.

The picture description task was based on Chartrand & Bargh (1999) with a more controlled time-course to allow averaging between trials. On each trial, one participant held a picture of a complex social scene, taken from the National Geographic website and printed on heavy card. The participant with the picture was the Leader for that trial, and had 30 seconds to describe the picture to the Follower, while the follower could not speak (monologue period). When a beep signalled the start of the 60 second dialogue phase, the follower was allowed to ask questions and the dyad could converse freely about the picture. Audible beeps indicated the start of the trial, the start of the monologue section and the end of the trial. A timer on the projector screen also counted down the time left in the monologue and dialogue sections. Thus, each picture description trial lasted 90 seconds with a fixed time structure in all trials. Each dyad completed 16 trials, alternating between speaker and listener roles (Fig 1).

At the end of the study, participants individually completed a questionnaire about the quality of their interaction. The questionnaire data is not analysed or reported here. Finally, participants were asked to write down what they thought was the purpose of the study, and were debriefed and paid by the experimenter.

Analyses

In line with our pre-registered plan, we excluded data from dyads who met any of the following criteria:

1. Participants knew each other before the study
2. Motion data was not recorded due to technical failure of the equipment or task software
3. Motion sensor(s) moved or fell off during the study
4. More than 50% of their data is missing or not suitable for wavelet analysis

Data were excluded from one dyad based on criteria 1 (familiarity), two dyads on criteria 3 (sensors fell off) and two dyads on criteria 4 (missing data). When carrying out the analyses, we also excluded individual trials if wavelet analysis could not be performed (e.g. this could happen if there is a very large jerk or jump in the motion data).

Pre-processing

Raw data was recorded as x-y-z coordinates and yaw-pitch-roll signals from 4 sensors – a head sensor and torso sensor of each participant, giving 24 channels of data at 60Hz for each 90 second trial. We trimmed data by discarding the first 100 time points (1.7s of the trial), all time points after the 5,250th point (87.5s into the trial; note that we originally specified 5300 in our pre-registration but some trials were shorter than 5300 data points). This removes irregularities at the start / end of trials and ensures that all trials are the same length. We corrected for circularity in the rotation data, to deal with cases where a change in orientation from -355° to 5° appears to Matlab like a large change rather than only a 10° movement past zero. To correct for small inaccuracies in timing, we resampled the data using a polyphase anti-aliasing filter (using the Matlab resample function). This was to ensure that we had a uniform fixed sample rate that exactly matched the number of expected data points. Each data channel was then de-trended by subtracting the mean value. Finally we applied a 7th order Butterworth low pass filter with cut-off frequency of 30 Hz to reduce noise.

Of the 24 data channels collected, the final analysis focused primarily on head pitch because this was the most informative signal in our pilot analysis. Other signals are presented in supplementary information. We conducted several different analyses, which we describe in relation to the figures.

Wavelet coherence analysis (Figure 2)

The raw data signals for the X and Y participants (Figure 2A&B) were subject to a wavelet transform and then the cross-wavelet coherence was calculated, using the Matlab toolbox from Grinsted et al. (2004) with default parameters (Morlet wavelet, w=6). To prevent edge effects from the start and stop times of the interaction (as well as the times surrounding the prompted switch from monologue to dialogue), we removed all coherence results falling outside the so-called 'cone of influence' (COI) (i.e. the pale areas in the corners of Fig 2C,D and E). In a very small minority of trials (mainly those with a single very jerky movement), the wavelet toolbox is unable to calculate the wavelet transform. Such trials were excluded from all analyses, and reported as missing data.

Next, we averaged the cross-wavelet coherence over the time course of each trial to obtain a simple measure of the frequency of coherence without regard to timing within a trial (Figure 2F). We truncated the wavelet output to frequencies from 0.2 – 8 Hz, resulting in a 1 x 89 vector of coherence values. This is a smaller frequency range that our original plan, because we found that data above 16 Hz could have been contaminated by dropped frames (affecting 2.7% of data points). Also, there was not enough data below 0.2 Hz (5 s period) to give a meaningful result, because with the COI removed, less than 10 seconds of useable monologue data remains below 0.2 Hz. Thus, our final output from the wavelet analysis is a coherence vector for each trial of each dyad, giving values of interpersonal coherence across frequencies from 0.2-0.8Hz.

Comparison to pseudo-pairs (Figure 3)

To fully quantify the levels of interpersonal coherence we observe in our data, we need to compare this to a null dataset where coherences is not present. We can do this by calculating the coherence present in pseudo-trials, where data from two different interactions are entered into the algorithm as if they were the X & Y participants (Bernieri & Rosenthal, 1991; Fujiwara & Daibo, 2016). Previous studies using this approach created pseudo trials by mixing datasets from different participants. We adopted a stricter approach where we create pseudo-trials using data from trials of the same dyad with the same leader-follower assignment. For example, we match X-trial 1 to Y-trial 3 where X is Leader and Y is Follower in both cases. Thus, our pseudo trials had the same general movement characteristics (e.g. overall signal power) as our real trials, and differed only in that the real trials represent a genuine live social interaction. We carried out wavelet analysis on the pseudo dataset using the same pipeline as above, and thus calculated the overall coherence values for all possible pseudo-trials of each dyad.

To implement group-level comparisons, we compare each dyad's real trial coherence to that dyad's pseudo trial coherence, using paired t-tests for each frequency bin. We correct for multiple comparisons using FDR (Benjamini & Hochberg, 1995) and plot the results both in terms of mean coherence levels and effect sizes for head pitch (Fig 3A and D), head roll (Fig 3B and E) and head yaw (Fig 3C and F).

Comparison of monologue and dialogue data (Figure 4)

The analysis described above used the full-trial data, with both monologue and dialogue analysed together. To explore potential differences between monologue and dialogue, we repeated the wavelet coherence analyses with the data separated into these two trial sections. Results are shown in Figure 4A, B, E and F. In addition, we used a basic analysis of power spectral density (PSD) over the whole trial (without breaking the data into wavelets) to check for global differences in signal power between participants and trial sections. For each participant and trial, we calculated the PSD using Matlab's pwelch function. Then we averaged data according to the participant's role as Leader or Follower in that trial. This allowed us to determine if there were overall differences in participants' movement behaviour between the leader / follower roles and monologue / dialogue sections (Figure 4, C, D, G and H).

Fast-nod detector (Figure 5)

Our wavelet analysis highlighted the 2.6-6.5Hz frequency range as an interesting band where one participant might engage in a fast-nodding behaviour while the other does not show coherent head motion. To explore this, we developed a simple 'fast nod detector' and test which participant shows fast-nods and when. Fast nods are defined in this work as head-pitch movements with a dominant frequency within the wider range of 1.5 to 8 Hz. This range is selected (rather than the significant findings of 2.6 to 6.5 Hz) to account for the approximate nature of our detector, and covers the range of frequency bands in Fig 3D with an effect size less than zero. One simple way of automatically detecting these frequencies is by thresholding on an estimate of the dominant frequency obtained from the zero-crossing rate (ZC). The ZC algorithm works by counting the number of times a signal crosses the zero (or window mean) within a given window, and is implemented here as follows:

1. slide a 2 second window across the pitch data,
2. high-pass filter by removing the window mean,
3. count the number of times the signal makes a zero crossing in one second (ZC),
4. calculate the zero-crossing rate as an approximation of the frequency (ZC/2 approximates frequency in Hz),
5. select those time windows where the approximate frequency falls within the wider range of 'fast nods' (between 1.5 and 8 Hz).

Dominant frequency can be computed by other means, like Fast Fourier Transform (FFT), however the ZC method has the advantage of simplicity and ease of implementation for future real-time execution. (One of the planned follow-on projects from this work is to implement an interactive virtual agent that can detect and respond to the head nods of human interlocutors in real time.)

The output of the fast-nod detector is a vector of 0/1 for each time point in each trial marking the presence / absence of a fast nod. Examples are shown in Figure 5A and B. Averaging this vector for each participant and trial gives an estimate of the rate of fast-nodding for that trial, and a paired t-test was used to compare rates between trials where that participant had a Leader role and those with a Follower role (Figure 5C). In addition, a speech-detector which thresholded the audio data was used to mark each timepoint in the data as ‘X speaking’, ‘Y speaking’ ‘Both speaking’ or ‘neither speaking’. Note that audio quality was too low to detect who was speaking in 2 dyads, so the sample size for this analysis is n=24 dyads. We used this to calculate the rate of fast nods for each participant when speaking and when not speaking, and then used a paired t-test to compare fast-nodding rates between Listening and Speaking phases within trials (Fig 5D). We can characterise the performance of the detector in terms of precision and recall. Precision measures the proportion of nods that occur during listening/following against all detected nods. Recall measures the proportion of nods that occur during listening/following against all periods marked as listening/following (Fawcett, 2003)

Exploring mimicry behaviour with phase – frequency histograms (Figure 6)

The wavelet coherence analysis revealed interpersonal coherence at low frequencies in our data. To explore the time-lags involved in more detail, we first used a cross-correlation analysis on our raw head-pitch data. We calculated the time-series cross-correlations between two participants for real trials and for pseudo-trials using Matlab’s xcorr function. This compares the correlation of two time-series for a range of temporal lags (in this case -4 to 4 seconds). We then averaged over trials and dyads, and used paired t-tests with FDR correction to contrast real and pseudo trial data (Fig 6 A and D).

When calculating wavelet coherence, it is possible to obtain information on both the coherence level and the phase difference between the two signals. Phase can only be meaningfully interpreted when there is positive coherence – that is, two signals are active in the same frequency range. For this reason we only store phase data when coherence meets a minimum threshold; here we choose a mid-range threshold of 0.5. For every dyad and trial, we calculated the phase difference between Leader and Follower (Fig 2G), and then thresholded this image to show only data points with a coherence over 0.5. For each frequency band, we then counted the number of supra-threshold points falling into each of 24 phase bins from -180^{o} to 180^{o}. This collapses the data over time and reveals the distribution of phases, and we can plot this data as a phase-frequency histogram (Fig 2H). We then average phase-frequency histograms over all trials and all dyads for both real trials and pseudo trials (Fig 6B and E). The difference between phase-frequency plots for real and pseudo trials (Fig 6C) reveals the frequency bands at which participants are in phase with a specific lag (yellow areas) and where less data is present than chance (blue areas). Thresholds on this map were created with paired-sample t-tests.

Modelling the phase-frequency histograms (Figure 7)

We aimed to test if the phase-frequency relationship seen in the lower part of Fig 6B (repeated in Fig 7A) is generated by a constant-phase mechanism or a constant-lag mechanism. To do this, we built two simple generative models, one for each mechanism. The constant-phase model had two parameters – phase lag and variability – and was modelled as a Gaussian distribution of phases about a fixed mean (Fig 7C). The constant lag model also had two parameters – time lag and variability – and was modelled by sampling individual trials, offsetting the 'X participant' movement by the time lag relative to the Y participant, then calculating the wavelet coherence and phase-frequency histogram for that sample trial. This process was repeated for 416 iterations (because the original dataset had 16x26 = 416 trials) using time-lags drawn from a Gaussian with mean & variability specified by the model parameters. The results of the 416 iterations were averaged to give the final result (Fig 7D). For each model, we used Matlab's fminsearch function to find the parameter values which gave the best fit between the model and the data, that is, to minimise the root-mean-squared error (RMSE) between the generated data (Fig7C or D) and the original data (Fig 7A). The model outputs shown in Fig 7 C and D represent the model using the optimal parameters. Comparing the RMSE values for the two models shows that the constant-lag model has a better fit to the data (Fig 7B). This implies that the cognitive mechanisms generating mimicry of head nods act with a constant lag of around 0.588 msec.

Data availability

Anonymised data & analysis code will be uploaded to OSF when this manuscript is accepted.

Acknowledgments

AH is supported by the ERC grant no 313398 INTERACT and by the Leverhulme Trust, JH was funded by an MoD-DSTL PhD studentship, JAW is funded by the Leverhulme Trust

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*

(Methodological). WileyRoyal Statistical Society. https://doi.org/10.2307/2346101

Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior*, *12*(2), 120–138. https://doi.org/10.1007/BF00986930

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. https://doi.org/10.1177/0146167294203008

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rim (Eds.), *Fundamentals of nonverbal behavior* (pp. 401–432). Paris, France: Editions de la Maison des Sciences de l'Homme.

Cappella, J. N. (1981). Mutual influence in expressive behavior: Adult–adult and infant–adult dyadic interaction. *Psychological Bulletin*, *89*(1), 101–132. https://doi.org/10.1037/0033-2909.89.1.101

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The Natural Statistics of Audiovisual Speech. *PLOS Comput Biol*, *5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Clark, H. (1996). *Using Language*. Cambridge University Press. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=b8bLCgAAQBAJ&oi=fnd&pg=PP1&dq=herb+clark+using+language&ots=iPi3Rd64VE&sig=AQYa9UfZaeIzE2TJd6p1jb2aRfU#v=onepage&q=herb clark using language&f=false

Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *The Journal of Nervous and Mental Disease*, *143*(4), 338–347.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, *58*(2), 342–353. https://doi.org/10.1037/0022-3514.58.2.342

Fawcett, T. (2003). ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. Retrieved from http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf?

Feese, S., Arnrich, B., Tröster, G., Meyer, B., & Jonas, K. (2011). Detecting posture mirroring in social interactions with wearable sensors (pp. 119–120).

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. https://doi.org/10.3758/BF03208827

Fujiwara, K., & Daibo, I. (2016). Evaluating Interpersonal Synchrony: Wavelet Transform Toward an Unstructured Conversation. *Frontiers in Psychology*, *7*, 516. https://doi.org/10.3389/fpsyg.2016.00516

Grammer, K., Kruck, K. B., & Magnusson, M. S. (1998). The courtship dance: patterns of nonverbal synchronization in opposite-sex encounters. *Journal of Nonverbal Behavior*, *22*(1), 3–29. https://doi.org/10.1023/A:1022986608835

Gunnery, S. D., Hall, J. A., & Ruben, M. A. (2013). The Deliberate Duchenne Smile: Individual Differences in Expressive Control. *Journal of Nonverbal Behavior*, *37*(1), 29–41. https://doi.org/10.1007/s10919-012-0139-4

Hadar, U., Steiner, T. J., Grant, E. C., & Rose, F. C. (1983a). Head Movement Correlates of Juncture and Stress at Sentence Level. *Language and Speech*, *26*(2), 117–129. https://doi.org/10.1177/002383098302600202

Hadar, U., Steiner, T. J., Grant, E. C., & Rose, F. C. (1983b). Kinematics of head movements accompanying speech during conversation. *Human Movement Science*, *2*(1–2), 35–46. https://doi.org/10.1016/0167-9457(83)90004-0

Hale, J. (2017). Using novel methods to examine the role of mimicry in trust and rapport. *Doctoral Thesis, UCL (University College London).* . Retrieved from http://discovery.ucl.ac.uk/1560499/

Hale, J., & Hamilton, A. (2016). Get on my wavelength: The effects of prosocial priming on interpersonal coherence measured with high-resolution motion capture. https://doi.org/none

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, *137*(3), 463–483. https://doi.org/10.1037/a0022288

Issartel, J., Bardainne, T., Gaillot, P., & Marin, L. (2014). The relevance of the cross-wavelet transform in the analysis of human interaction - a tutorial. *Frontiers in Psychology*, *5*(January), 1566. https://doi.org/10.3389/fpsyg.2014.01566

Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta Psychologica*, *32*, 101–125. https://doi.org/10.1016/0001-6918(70)90094-6

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology (2006)*, *63*(11), 2220–2230. https://doi.org/10.1080/17470218.2010.497843

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjodt, U., Jegindo, E.-M., Wallot, S., … Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences*, *108*(20), 8514–8519. https://doi.org/10.1073/pnas.1016955108

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., Maciver, M. A., & Poeppel, D. (2017). Neuron Perspective Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*, 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). THE CHAMELEON EFFECT AS SOCIAL GLUE : EVIDENCE FOR THE EVOLUTIONARY SIGNIFICANCE OF NONCONSCIOUS MIMICRY. *Journal of Nonverbal Behavior*, *27*(3), 145–162. https://doi.org/10.1023/1025389814290

Lavan, N., Scott, S. K., & McGettigan, C. (2016). Laugh Like You Mean It: Authenticity Modulates Acoustic, Physiological and Perceptual Properties of Laughter. *Journal of Nonverbal Behavior*, *40*(2), 133–149. https://doi.org/10.1007/s10919-015-0222-8

Leander, N. P., Chartrand, T. L., & Bargh, J. a J. A. (2012). You give me the chills: embodied reactions to inappropriate amounts of behavioral mimicry. *Psychological Science*, *23*(7), 772–779. https://doi.org/10.1177/0956797611434535

Morrill, R. J., Paukner, A., Ferrari, P. F., & Ghazanfar, A. A. (2012). Monkey lipsmacking develops like the human speech rhythm. *Developmental Science*, *15*(4), 557–568. https://doi.org/10.1111/j.1467-7687.2012.01149.x

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual Prosody and Speech Intelligibility. *Psychological Science*, *15*(2), 133–137. https://doi.org/10.1111/j.0963-7214.2004.01502010.x

Noorden, L. van, & Moelants, D. (1999). Resonance in the Perception of Musical Pulse. *Journal of New Music Research*, *28*(1), 43–66. https://doi.org/10.1076/jnmr.28.1.43.3122

Ohala, J. J. (1975). The temporal regulation of speech. In *Auditory Analysis and Perception of Speech*.

Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., & Kelso, J. a S. (2008). Social coordination dynamics: measuring human bonding. *Social Neuroscience*, *3*(2), 178–192. https://doi.org/10.1080/17470910701563392

Paxton, A., & Dale, R. (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods*, *45*(2), 329–343. https://doi.org/10.3758/s13428-012-0249-2

Pellegrini, R., & Ciceri, M. R. (2012). Listening to and mimicking respiration: Understanding and synchronizing joint actions. *Review of Psychology*, *19*(1), 17–27.

Pentland, A. (2008). *Honest Signals*. MIT Press.

Poppe, R., Van Der Zee, S., Heylen, D. K. J., & Taylor, P. J. (2013). AMAB: Automated measurement and analysis of body motion. *Behavior Research Methods*, *46*(3), 625–633. https://doi.org/10.3758/s13428-013-0398-y

Ramseyer, F., & Tschacher, W. (2010). Nonverbal Synchrony or Random Coincidence? How to Tell the Difference. In A. Esposito, N. Campbell, C. Vogel, A. Hussain, & A. Nijholt (Eds.), *Development of Multimodal Interfaces: Active Listening and Synchrony* (pp. 182–196). Springer Berlin Heidelberg.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, *79*(3), 284–295. https://doi.org/10.1037/a0023419

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schmidt, R. C., Morr, S., Fitzpatrick, P., & Richardson, M. J. (2012). Measuring the Dynamics of Interactional Synchrony. *Journal of Nonverbal Behavior*, *36*(4), 263–279. https://doi.org/10.1007/s10919-012-0138-5

Schmidt, R. C., Nie, L., Franco, A., & Richardson, M. J. (2014). Bodily synchronization underlying joke telling. *Frontiers in Human Neuroscience*, *8*, 633. https://doi.org/10.3389/fnhum.2014.00633

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326–332. https://doi.org/10.1037/0096-1523.29.2.326

Stel, M., van Dijk, E., & Olivier, E. (2009). You want to know the truth? Then don't mimic! *Psychological Science*, *20*(6), 693–699. https://doi.org/10.1111/j.1467-9280.2009.02350.x

Tomasello, M. (1995). *Joint attention as social cognition*. (C. Moore & P. J. Dunham, Eds.). Hillsdale, NJ: Erlbaum.

van Baaren, R. B., Decety, J., Dijksterhuis, A., van der Leij, A., & van Leeuwen, M. L. (2009). Being imitated: Consequences of nonconsciously showing empathy. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 31–42). Cambridge, MA, US: MIT Press.

Walton, A. E., Richardson, M. J., Langland-Hassan, P., & Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00313

Warner, R. M. (1996). Coordinated cycles in behavior and physiology during face-to-face social interactions. In J. H. Watt & C. A. VanLear (Eds.), *Dynamic patterns in communication processes* (pp. 327–352). Thousand Oaks, CA, US: Sage Publications, Inc.