Co-ordination of Head Nods: Asymmetries between Speakers and Listeners

Leshao Zhang
Human Interaction Lab
Cognitive Science Research Group
Queen Mary University of London
leshao.zhang@qmul.ac.uk

Patrick G.T. Healey
Human Interaction Lab
Cognitive Science Research Group
Queen Mary University of London
p.healey@qmul.ac.uk

Abstract

Previous research suggests that if people unconsciously mimic their interaction partner's movement, they gain social influence. We compare the effectiveness of speakers that mimic listeners' head nods with speakers that use natural nods in a special customised virtual environment. The results suggest that listeners agreed more with mimicking speakers than natural speakers. However, there are also asymmetries in speaker-listener nodding in the high and low-frequency domain. Listeners nod significantly more than speakers in the high-frequency domain. This asymmetry may be an important factor in coordination. We conclude that speaker and listener nods have both different form and different functions.

1 Introduction

There is significant interest in the coordination of speaker and listener behaviour in conversation, especially mimicry of form and/or temporal synchronisation of behaviour. Previous research has suggested that people automatically mimic each others' movements and behaviours unconsciously during interaction (Chartrand and Bargh, 1999), usually within a short window of time of between three to five seconds. It is claimed that this can prompt changes in individuals' cognitive processing style, altering performance on tests of ability and creativity and shifting preferences for consumer products as well as improving liking, empathy, affiliation, increasing help behaviour and reducing prejudice between interactants (Chartrand and Lakin, 2013). Based on this idea, Bailenson and Yee conducted the "Digital Chameleons" (2005) study. They created a virtual speaker automatically mimic the listener's head nods and suggested that the mimicking agent was more persuasive than the nonmimicker. However, these effects have not been consistently replicated (Riek et al., 2010; Hale and Hamilton, 2016; Zhang and Healey, 2018).

In this paper, we investigate mimicry effects in more detail by comparing natural, mimicked, acted or 'canned' (i.e. non-interactive) playback of nodding behaviour in dialogue. These experimental manipulations are achieved through the use of a special customised Immersive Virtual Environment (IVE) which supported multiple people real-time interaction. For each of these manipulations, we explore the dynamics of the joint head movements both inside the virtual environment i.e. what the participants see and respond to and compare this with the coordination of their actual nodding behaviours.

2 Background

2.1 Nonverbal Studies with Immersive Virtual Environments

Immersive virtual environments (IVEs) have provided new ways to experiment with nonverbal interaction (Blascovich et al., 2002; Healey et al., 2009; Bailenson et al., 2001). In face-to-face interaction studies, it is difficult to introduce experimentally controlled manipulations of nonverbal behaviours. In principle, IVEs enable control of all aspects of participant's non-verbal behaviour (Bailenson et al., 2001). They also provide researchers with access to all participant's motion data, including all visible movements, gaze, and gestures (Blascovich et al., 2002). This 'panoptic' capability allows for subsequent analysis of all behaviours from any arbitrary viewpoint, something that is impossible with video.

2.2 Digital Chameleons

The "Digital Chameleons" study (Bailenson and Yee, 2005) illustrates the potential of IVEs. Bailenson and Yee compared the persuasiveness of a virtual agent which automatically mimics a listener's head nods at a 4 seconds delay with an agent which reproduced a previous listener's head nods (so playback of naturalistic head nods but random with respect to what is being said in the interaction). They found evidence that a mimicking agent is more persuasive than the playback condition when delivering a message to students that they should always carry their ID card.

Similar studies were repeated over recent years. Researchers either found the effects of the "Digital Chameleons" (Bailenson and Yee, 2007; Verberne et al., 2013; Stevens et al., 2016) or could not consistently replicate the result (Riek et al., 2010; Hale and Hamilton, 2016; Zhang and Healey, 2018). This suggested that we might not have enough understanding of the speaker-listener head-nodding coordination.

2.3 Head Nods

Head nods are an important conversational signal. They are the most frequent head movement behaviour among shakes and changes of angle/orientation, etc (Włodarczak et al., 2012; Ishi et al., 2014). One possible reason for the mixed evidence on head-nodding coordination is the potential for different kinds of nod with different frequencies.

Hader et al. (1983) distinguishes three different head nods by frequency: 1) slow head nods between 0.2-1.8 Hz 2) ordinary head nods between 1.8-3.7 Hz and 3) rapid head nods above 3.7 Hz. They also suggest that listeners mainly use ordinary head nods to signal 'YES', rapid head nods for synchrony and slow/ordinary nods for other tasks. Other definitions of head nods by speed have been used. For example, Hale et al. (2018) define slow head nods as between 0.2-1.1 Hz, fast head nods between 2.6-6.5 Hz and found that listeners produce more fast head nods than speakers.

Head nods also serve different functions for listeners and speakers, e.g., listeners use "back channel" nods to signal their agreement, interests or impatience or to synchronise with a speaker's head nods (Hadar et al., 1985); while speakers may nod to seek or check agreement, to signal continued speaking, to express emphasis or as 'beat' gestures that accompany the rhythmic aspects of speech (Heylen, 2005). Listener head nods are also primarily concurrent with the speaker's turn. Healey et al. (2013) showed that speakers nod more than primary addressees and that this relationship varies depending on how fluent the speaker's performance is.

2.4 Cross Recurrence Quantification Analysis

Analysis of the coordination of speaker and listener head nods requires methods that can find coordinated patterns in time-series over a variety of temporal intervals.

Recurrence Quantification Analysis (RQA) (Webber Jr and Zbilut, 2005) is a nonlinear time-series analysis method for the analysis of chaotic systems. Cross Recurrence Quantification Analysis (xRQA) is RQA applied to two independent time-series, e.g., two participants and finds the degree of match between the two time-series at different temporal offsets. So, for example, it can detect if one person's nods are systematically repeated by another person. xRQA has been widely used in the analysis the coordination of the interactants in a conversation (Richardson and Dale, 2005; Dale and Spivey, 2006; Richardson et al., 2008).

xRQA reconstructs two one-dimensional time-series data to pairs of points in a higher Embedding Dimension phase space (Takens, 1981) using Time-Lagged copies. It calculates the distances between the reconstructed pairs of points. The points pair that fall within a specified distance (Radius) are considered to be recurrent. The recurrent points are visualised with Recurrence Plots (RPs) that show the overall amount of repetition of (%REC), the longest sequence of repeated behaviours (LMAX) and the predictability or determinism (%DET) of one sequence from another. More specifically, %REC is the percentage of recurrent points in the RP. It indexes how much the two time-series are repeated. LMAX is the length of the longest diagonal line segment in the RP. It indexes the coupling strength of the two time series. %DET is the percentage of recurrent points falls on diagonal lines. It shows how much one time-series is predictable from another.

3 Current Study

To investigate the coordination of speaker-listener head nods, we used a customised IVE (Figure 1) that supports multiple participants' real-time interaction. Participants interact through virtual avatars. An optical motion capture system (Vicon)

Figure 1: The IVE in the Listener's View

Figure 2: Two Participants Were Doing the Experiment

captures participant's body movements in real-time, and this drives the movement of the avatars inside the IVE. Eye and lip movements are not captured so an algorithm is used to generate naturalistic eye movements and vocal amplitude is used to drive lip movements (previous research suggests participant's find the animation broadly realistic (Zhang and Healey, 2018)). In this study we used an asymmetrical setting for the speaker-listener interaction: the listener is immersed into the IVE and sees the speaker as a virtual character while the speaker is not immersed but is in the same physical room as the listener (see Figure 2).

3.1 Procedure

One participant acts as a listener and, in the appropriate conditions, a second participant acts as the speaker. In each experiment trial, participants wear the marker suits for motion tracking after an introduction. The listener also wears the Oculus Rift HMD and interacts with the virtual representation of the speaker (avatar) in the IVE. Following Bailenson et al. (2005), the speaker is asked to deliver a short pre-written speech about student ID card regulations to the listener. The speaker faces the listener and can see their body movements but cannot make eye-contact (Figure 2). The monologue is about 2-3 minutes long. After the monologue, the listener is asked to fill an online questionnaire on a computer.

In the experiment, the virtual speaker's head nods are manipulated according to the assigned condition:

1. Mimic – the virtual speaker's head nods are exact mimics of the listener's head nods but at a 4s delay.

2. Playback – the virtual speaker's head nods are an exact replay of the nods of the previous listener's head nods.

3. Natural – the virtual speaker's head nods are an exact mapping of the real speaker's head nods.

4. Recording – the virtual speaker's full body movements are an exact replay of a pre-recorded animation of a speaker/actor.

The Mimic, Playback and Natural conditions were assigned in rotation while the Recording condition was applied whenever we had only one participant in an experimental trial.

3.2 Measures

The analysis is organised into two sections. First, subjective assessments of the effectiveness of the speaker. Second, the patterns of head-nodding behaviour for the virtual and real speaker-listener pairs as determined from the motion capture data.

3.2.1 Effectiveness of the Speaker

We did exactly the same measurement for the speaker's effectiveness as Bailenson and Yee did in the "Digital Chameleons" study. The effectiveness of the speaker was measured by listener ratings on a self-report questionnaire. Speaker effectiveness is assessed by 4 items about agreement (agreement, valuable, workable, needed of the student ID card regulation delivered by the speaker), 13 items (friendly, likeable, honest, competent, warm, informed, credible, modest, approachable, interesting, trustworthy, sincere and overall) on impressions of the speaker, and 8 items (enjoy, acceptable, isolating, attractive, comfortable, cooperative, self-conscious or embarrassed and overall) on the virtual speaker's social presence; with Likert scale range from 1 strongly disagree to 7 strongly agree. Based on our previous research, we made our null hypothesis:

H0 The effectiveness of the speaker does not differ across conditions.

3.2.2 Amount of Head Nods

The body movement data was recorded as the joint orientation time-series in degrees at 60 Hz. With the recorded head movement time-series data, we tested the difference of the number of head nods between the speaker and listener with the paired t-test in the frequency range 0-8 Hz. Peaks in the head-nodding time-series were treated as the point that the participant changed the direction of head movement and counted as a nod. The total amount of head nods was counted as the number of peaks in the head-nodding time-series data. A low pass filter was used on the time-series data with the cut-off frequency set to increase slowly from 0 to 8 Hz in the resolution of 0.1 Hz. Building on previous work our initial hypothesis was that:

H1 Speakers nod more than listeners in all the conditions.

3.2.3 Head-Nodding Coordination

The coordination of speaker-listener head-nodding was tested using the xRQA method. We calculated a baseline chance coordination of the speaker-listener nods by doing xRQA with randomly paired speaker's and listener's from the Natural condition. We compared the head-nodding coordination in each condition as well as the chance level coordination for both the virtual and real speaker-listener pair. Given the assumption that non-verbal communication is coordinated in actual interactions, our second hypothesis is:

H2 Coordination of the speaker-listener head nods in all conditions is higher than chance.

3.3 Pairing Participants

Instead of running separate pairs of participants in each trial, we applied a shifted overlay participant arrangement. Each participant took part in two conditions. As shown in Figure 3, participants were asked to act first as a listener, then as a speaker. In each experimental trial, we had a previous participant as the speaker and a current participant as the listener. This setting ensured that before every experiment trial, the speaker has already been in the virtual environment and heard the message delivered by the previous speaker. Thus, the speaker would understand what the listener would see in the virtual world and be familiar with the message they would need to deliver to the listener. In the case of only one participant presented in the experiment, e.g. the very first experiment trial or one participant was not showing up, we replaced the speaker with an animated virtual agent to deliver the message which corresponded to the recording condition.

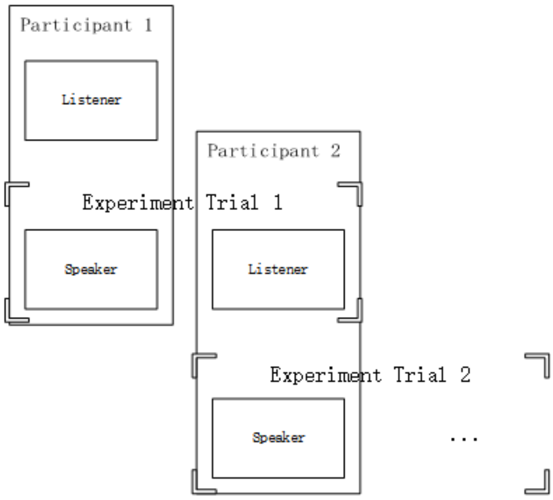


Figure 3: The Procedure Flow of Experiment II

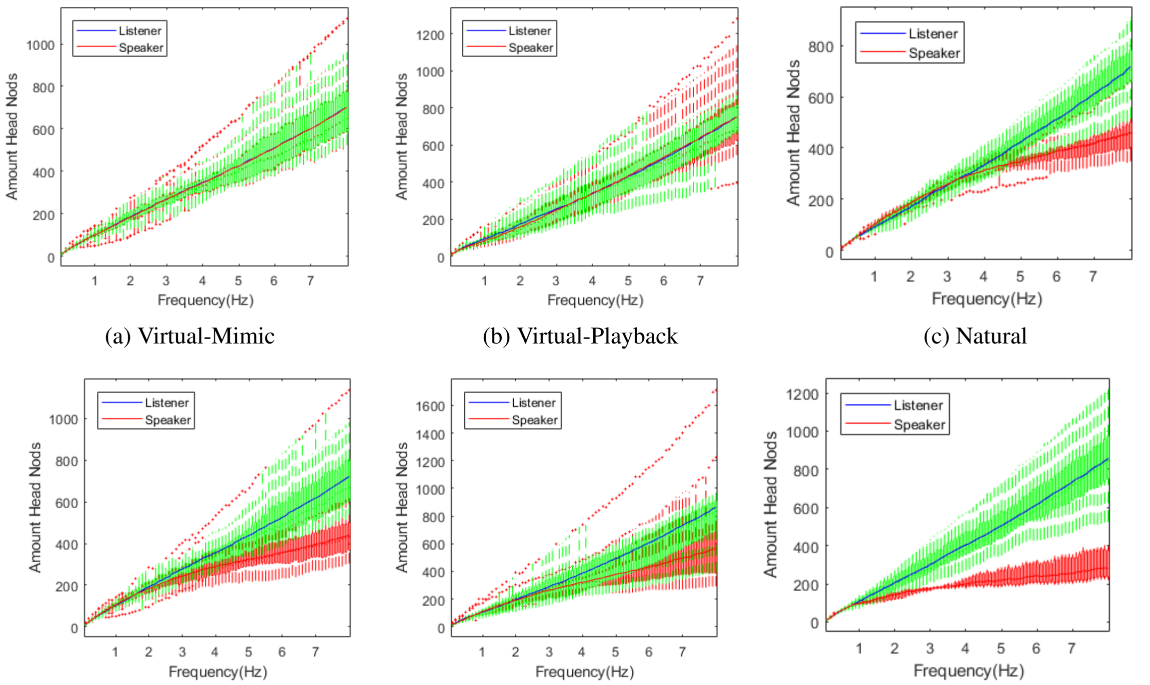
3.4 Participant

54 participants were recruited by email, posters and through a participant panel. Each participant received 10 pounds for their participation. The final sample consisted of 29 female and 25 male students between 18 to 33 (Mean=21.89, SD=3.45). None of the participants reported severe motor, auditive or visual disabilities/disorders.

3.5 Result

3.5.1 Effectiveness of the Speaker

We tested the agreement, impression, social presence and the overall effectiveness of the speaker with the Generalized Linear Mixed Model (GLMM) analysis with the fixed factors of experiment condition (Mimic, Playback, Natural, Recording). Subject, speaker/listener's gender and the rating of their relationship were included as random effects. The result suggested that the listener's agreement with the speaker is slightly higher in the Mimic condition than in the Natural condition, t_{50}=2.218, p=0.031; the listener's impression of the speaker is higher in the Mimicry condition than in Recorded condition, t_{50}=2.655, p=0.011; the social presence of the speaker is higher in the Playback condition than in the Recorded condition, t_{50}=2.870, p=0.006; the overall effectiveness of the speaker is higher in the Mimic condition than in the Recorded condition, t_{50}=2.491, p=0.016. No other significant effect was found.



(a) Virtual-Mimic (b) Virtual-Playback (c) Natural

(d) Reality-Mimic (e) Reality-Playback (f) Recording

Figure 4: Boxplots of the Cumulative Amount of Head Nods for the Virtual and Real Speaker-Listener Pair.

3.5.2 Amount of Head Nods

We counted the number of head nods for every pair of participants. Figure 4 shows the distribution of the number of head nods for the virtual and real speaker-listener pair with a series of boxes. The X-axis is the cutoff frequency of the low pass filter. The Y-axis is the number of head nods for the participants through a certain low pass filter. The boxes were taken in the resolution of 0.1 Hz.

We compared the mean difference of the number of head nods between the listener and speaker below the certain frequency with the paired t-test. The result suggested that for the virtual pair of speaker and listener, there was no significant difference of the number of head nods under the condition of mimic and playback with the exception that the listener has a significantly higher amount of head nods than the real speaker in the frequency range from 4-8 Hz. Moreover, in the natural condition, the listener nodded less in the frequency range between 0.7-1.5 Hz whereas nodded more in the frequency between 3-8 Hz than the speaker. In the recording condition, the listener nodded significantly more than the speaker beyond 1 Hz. Figure 5 and 6 shows the mean difference of the number of head nods between the speaker and listener (listener to speaker) for the virtual and real pair respectively. The red dots in the graph indicate the points are under the significant level of 0.05.

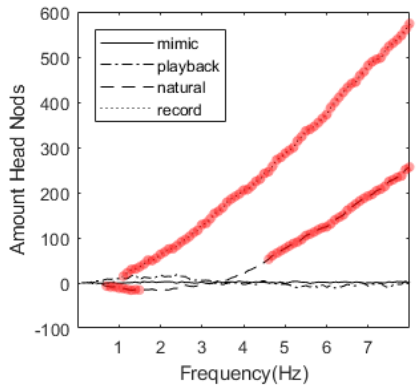


Figure 5: Cumulative Mean Difference of the Amount of Head Nods for the Virtual Listener-Speaker Pair.

3.5.3 Head-Nodding Coordination

xRQA was run for all the virtual and real interactional pairs with fixed parameters: Embedding Dimension=6, Time Lag=1, Radius=50, Non-normalised. The fixed parameters ensured that the parameters were kept as the controlled variables; the value of the parameters was picked to ensure no floor or ceiling effect for the xRQA outputs; not normalise the data to reduce the effect of non-movement. Figure 7 is the RP examples for the speaker-listener pair in each condition. The RPs in

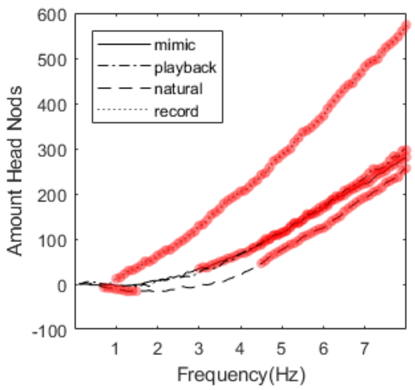


Figure 6: Cumulative Mean Difference of the Amount of Head Nods for the Real Listener-Speaker Pair.

mimic and playback condition were divided into virtual and real pair, whereas the RPs in natural and recording condition were not as they move the same in the virtual or real world. As we can see with the RPs, the virtual pairs with the mimic (Figure 7a) and playback (Figure 7b) condition were more coordinated (more dots in the RP) than the other conditions. They showed different coordinating patterns, e.g., there was a long diagonal line in the RPs of the mimic condition which was not seen in the RPs of the playback condition. The diagonal line has a tiny offset in Y-axis which indicated the 4s delay mimicry manipulation of the virtual speaker's head-nodding. The RPs of the recording (Figure 7f) condition showed the least coordination (least dots) of the speaker-listener pair. However, we cannot easily tell the difference between the RPs of the mimic (Figure 7d), playback (Figure 7e) and natural (Figure 7c) conditions with the real pairs.

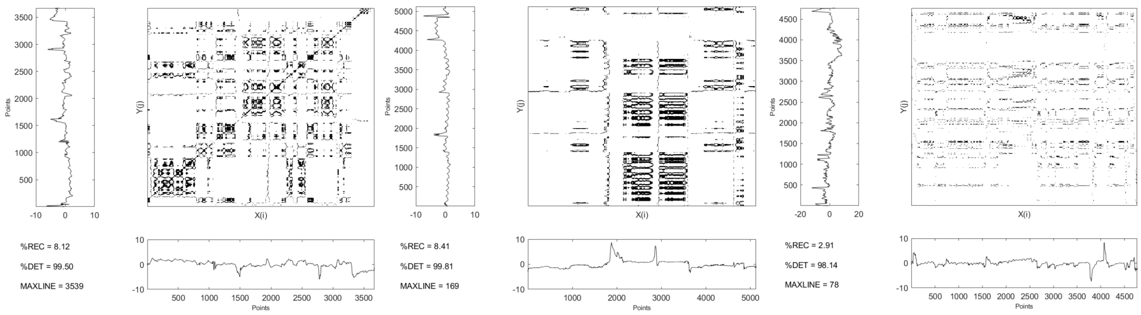
The quantification outputs of the xRQA calculated the %REC, LMAX and %DET for all the virtual and real speaker-listener pairs. Figure 8 is the boxplots for those xRQA outputs by condition. The horizontal red lines are the chance level of these measures with the 95% confidence interval. We tested the %REC, LMAX and %DET for virtual and real speaker-listener pairs between conditions. The result suggested there was a significant ($p<0.001$) difference between conditions on these items for the virtual and real speaker-listener pairs.

Games-Howell posthoc pairwise test suggested that: for the virtual speaker-listener pair, %REC was not significantly different from the chance level in the mimic, playback and natural condition, while it was significantly below the chance level in the recording condition, Mean Difference (MD)=2.72, $p<0.001$; with LMAX mimic was great than playback (MD=4588, $p<0.001$), playback was great than natural (MD=99.4, $p<0.005$), natural was at about chance level and great than recording (MD=32.5, $p<0.001$); %DET was above the chance level in the mimic (MD=2.75, $p<0.001$) and playback (MD=3.0, $p<0.001$) conditions, and below the chance level in the recording condition (MD=4.37, $p<0.001$), while not different from the chance level in the natural condition. For the real speaker-listener pair, %REC was below the chance level in the recording condition (MD=2.72, $p<0.001$) while no significant difference from the chance level in the mimic, playback and natural condition; LMAX was not reliably different from chance in the mimic and natural conditions, whereas it was above the chance level in the playback condition and below the chance level in the recording condition; %DET was not significantly different from the chance level in the mimic, playback and natural condition, while it was significantly below the chance level in the recording condition (MD=4.37, $p<0.001$).

4 Discussion

The results suggest that listeners may agree more with the speaker in the Mimic condition than in the Natural condition. Although this would indicate rejection of the null hypothesis H0, the evidence here is weak given the number of statistical comparisons made. There was also a difference in the effectiveness of the speaker when we manipulated its head movement behaviour. This was a surprise to us as we expected that there would be no difference in the speaker's effectiveness across all the conditions. Overall, the present study does not provide clear evidence for an effect of mimicry on agreement and persuasion but does indicate this might be worth pursuing in further work.

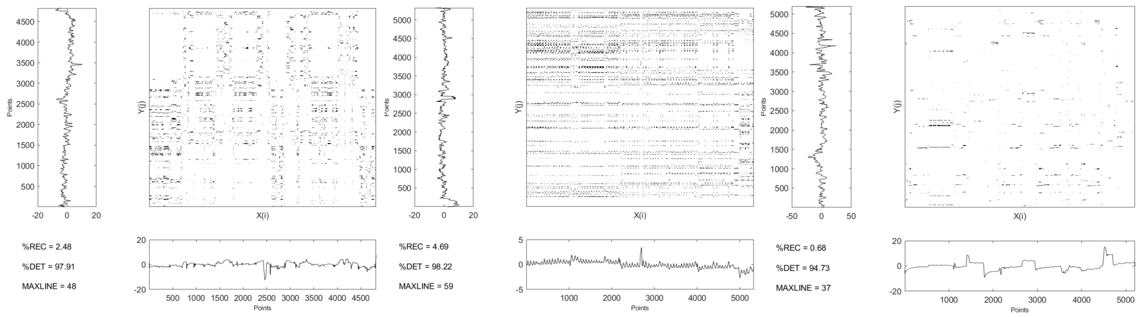
A much more salient and surprising finding is the distribution of head-nodding behaviour by the speaker and listener during the monologue. In terms of the number of head nods, the results show that listeners nodded significantly more in the high-frequency domain (above 3 Hz), and less in the low-frequency domain (between 0.7-1.5 Hz) in the Natural condition while no difference was observed in the other conditions. This suggests that we partly reject the hypothesis H1. In natural communication, speaker and listener nod dif-



(a) Mimic Condition for Virtual Pair

(b) Playback Condition for Virtual Pair

(c) Natural Condition

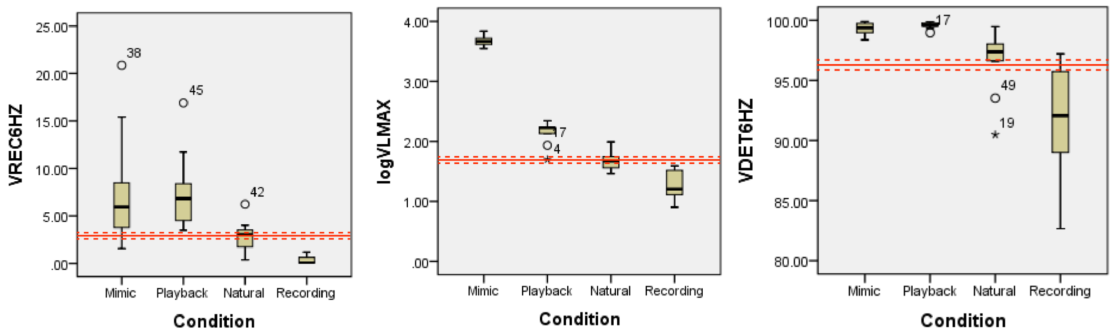


(d) Mimic Condition for Real Pair

(e) Playback Condition for Real Pair

(f) Recording Condition

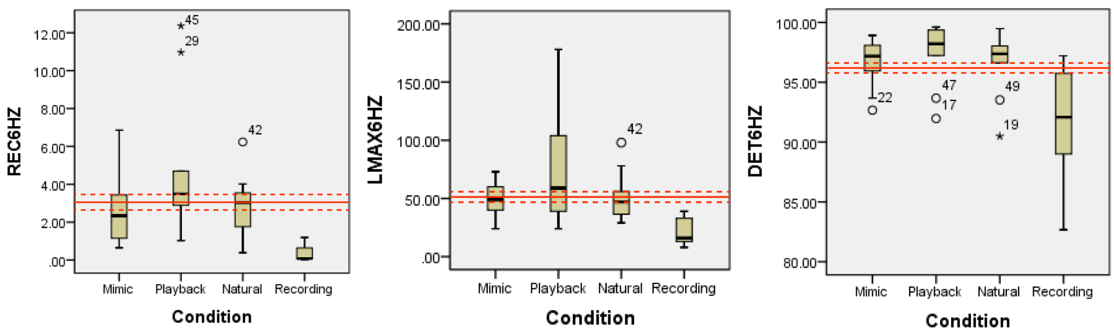
Figure 7: The Recurrence Plot for Speaker-Listener Pair in Each Condition



(a) %REC for Virtual Pair

(b) LMAX for Virtual Pair*

(c) %DET for Virtual Pair



(d) %REC for Real Pair

(e) LMAX for Real Pair

(f) %DET for Real Pair

Figure 8: Boxplots of xRQA Outputs for the Virtual and Real Speaker-Listener Pair. Red line is the mean value of random pair with 95% CI. *A logarithm to base 10 was applied to the LMAX for Virtual Pair to compress the scale as the value in mimic condition is extremely high due to the experimental manipulation.

ferently in the high and low frequency domain (cf. (Hale et al., 2018)). Moreover, Figure 4f indicated that the speaker in the recording condition nodded much less in the high-frequency domain than the speaker in the other conditions. This is despite the fact that people performing the monologue in the Recorded condition moved much more overall than any of the other speakers. This might be because in the absence of a real listener, speakers perform significantly fewer fast nods. If fast nods are listener specific behaviours they might be a key contribution to the reciprocal dynamics between speakers and listeners. In other words, using an actor to perform a communication with the absence of the real listener leads to a non-verbal performance that is very different from the natural behaviour of a speaker in a live interaction - even when it is a monologue.

We also tested the speaker-listener's head-nodding coordination by applying the one-way ANOVA to the xRQA outputs. The most obvious point about the results illustrated in Figure 8 is that coordination with the Recorded speaker is consistently well below our measure of chance. The primary reason for this is that the people who recorded their monologues moved much more than those who delivered or listened to them live. These movements rarely matched those of their listeners who were relatively still.

Interestingly, the results also show that speaker-listener head-nodding coordination is not different chance in the Natural condition. In these data head-nodding coordination only exceeds chance in the Mimic and Playback conditions in the virtual speaker-listener pairs and is not different from chance with the real speaker-listener pairs. This is unsurprising in the virtual mimicry case since the experimental manipulation guarantees that nods are mimicked. The above chance coordination in the virtual Playback case is more puzzling. One possible explanation is that it occurs because we are pairing the head movements of listeners with listeners. Since the results indicate that listener head movements have a different characteristic frequency, this makes chance similarity higher than it is for speaker-listener combinations. This suggests accepting the null hypothesis for H2 as well. Natural speaker-listener head-nodding is no more coordinated than we could expect by chance Recorded virtual speaker's head-nodding is significantly decoupled.

It is interesting to note that overall coordination of speaker-listener head-nodding is higher in the virtual world than in the real world with the mimic and playback conditions. The only difference between the two worlds is the speaker's head nods. In the virtual world, the speaker's head nods are taken from a listener, either from the listener themselves (Mimic condition) or from another listener (Playback condition), whereas in the real world, they are their actual head nods. Since listeners nod more than the speaker in the high-frequency domain, this could account for the elevated levels of virtual coordination. This is consistent with previous works (Hadar et al., 1985; Hale et al., 2018).

A potential limitation of the experimental approach used here is that the relation of the timing of head nods and vocal stress in the speech is not controlled. For example, Giorgolo and Verstraten (2008) suggest that temporally shifting the timing of hand gestures in the video away from its audio component create an anomalous feeling. Although only one participant (out of 54) reported a detachment of the head nods from the speech in debriefing, the effect of the correlation between the timing of speaker's head nods and the vocal stress in the speech is not clear in this work and needs further study.

5 Conclusion

The results suggest that in some circumstances speakers get more agreement by mimicking listener nodding behaviour. However, they also show that speaker and listener head nods are different in character. In the Natural interaction condition people do not coordinate their nodding behaviour more than would be expected by chance. The analysis of head-nodding behaviour suggests that this is because speakers nod more in the low-frequency domain and less in the high-frequency domain than the listener. The speaker-listener head-nodding coordination is above chance for the mimicking speaker, at chance for the natural speaker and below chance for an animated (recorded) virtual speaker. We also found that the fast nods are critical in the speaker-listener's coordination.

Acknowledgments

The work is supported by EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

References

Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2001. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 10(6):583–598.

Jeremy N. Bailenson and Nick Yee. 2005. Digital chameleons. *Society*, 16(10):814–819.

Jeremy N. Bailenson and Nick Yee. 2007. Virtual interpersonal touch and digital chameleons. *Journal of Nonverbal Behavior*, 31(4):225–242.

Jim Blascovich, Jack Loomis, Andrew C Beall, Kimberly R Swinth, Crystal L Hoyt, and Jeremy N Bailenson. 2002. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2):103–124.

Tanya L. Chartrand and John A. Bargh. 1999. The chameleon effect: The perceptionbehavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893–910.

Tanya L Chartrand and Jessica L Lakin. 2013. The antecedents and consequences of human behavioral mimicry. *Annual review of psychology*, 64:285–308.

Rick Dale and Michael J Spivey. 2006. Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, 56(3):391–430.

Gianluca Giorgolo and Frans A. J. Verstraten. 2008. Perception of speech-and-gesture integration. In *In Proceedings of the International Conference on Auditory-Visual Speech Processing*, pages 31–36.

U. Hadar, T.J. Steiner, E.C. Grant, and F. Clifford Rose. 1983. Head movement correlates of juncture and stress at sentence level. *Language and Speech*, 26(2):117–129. PMID: 6664179.

Uri Hadar, Timothy J Steiner, and F Clifford Rose. 1985. Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, 9(4):214–228.

Joanna Hale and Antonia F. De C. Hamilton. 2016. Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific Reports*, 6.

Joanna Hale, Jamie A Ward, Francesco Buccheri, Dominic Oliver, and Antonia Hamilton. 2018. Are you on my wavelength? interpersonal coordination in naturalistic conversations.

Patrick G T Healey, Mary Lavelle, Christine Howes, Stuart Battersby, and Rosemarie McCabe. 2013. How listeners respond to speaker's troubles. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35.

Patrick G.T. Healey, Chris Frauenberger, Marco Gillies, and Stuart Battersby. 2009. Experimenting with non-verbal interaction. In *8th international gesture workshop*.

Dirk Heylen. 2005. Challenges ahead: Head movements and other social acts in conversations. In *Proceedings of Joint Symposium on Virtual Social Agents*, pages 45–52.

Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. 2014. Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, 57:233–243.

Daniel C Richardson and Rick Dale. 2005. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, 29(6):1045–1060.

Michael J Richardson, Stacy Lopresti-Goodman, Marisa Mancini, Bruce Kay, and RC Schmidt. 2008. Comparing the attractor strength of intra- and interpersonal interlimb coordination using cross-recurrence analysis. *Neuroscience Letters*, 438(3):340–345.

Laurel D Riek, Philip C Paul, and Peter Robinson. 2010. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, 3(1-2):99–108.

Catherine J Stevens, Bronwyn Pinchbeck, Trent Lewis, Martin Luerssen, Darius Pfitzner, David MW Powers, Arman Abrahamyan, Yvonne Leung, and Guillaume Gibert. 2016. Mimicry and expressiveness of an eca in human-agent interaction: familiarity breeds content! *Computational cognitive science*, 2(1):1.

Floris Takens. 1981. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer.

Frank M. F. Verberne, Jaap Ham, Aditya Ponnada, and Cees J. H. Midden. 2013. Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust. In *Persuasive Technology*, pages 234–245, Berlin, Heidelberg. Springer Berlin Heidelberg.

Charles L Webber Jr and Joseph P Zbilut. 2005. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, pages 26–94.

Marcin Włodarczak, Hendrik Buschmeier, Zofia Malisz, Stefan Kopp, and Petra Wagner. 2012. Listener head gestures and verbal feedback expressions in a distraction task. In *Feedback Behaviors in Dialog*.

Leshao Zhang and Patrick G.T. Healey. 2018. Human, chameleon or nodding dog? In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pages 428–436. ACM.