Coordination of Nods in Dialogue

Leshao Zhang

Submitted in partial fulfillment of the requirement of
the Degree of Doctor of Philosophy

Media and Arts Technology
School of Electronic Engineering and Computer Science
Queen Mary University of London

Feb. 2020

Abstract

Behavioral mimicry has been claimed to be a nonconscious behavior that evokes prosocial effects — liking, trust, empathy, persuasiveness — between interaction partners. Recently Intelligent Virtual Agents (IVAs) and Immersive Virtual Environments (IVEs) have provided rich new possibilities for nonverbal behavior studies such as mimicry studies. One of the best known effects is the "Digital Chameleons" in which an IVA appears to be more persuasive if it automatically mimics a listener's head nods. However, this effect has not been consistently replicated. This thesis explores the basis of the "chameleon effects" using a customized IVE integrated with full-body motion capture system that support real-time behavior manipulation in the IVE. Two replications exploring the effectiveness of the virtual speaker and head nodding behavior of interaction partners in the agent-listener interaction and avatar-listener interaction by manipulating the virtual speaker's head nods and provide mixed results. The first experiment fails to replicate the original finding of mimicry leading to higher ratings of an agent's effectiveness. The second experiment shows a higher rating for agreement with a mimicking avatar. Overall, an avatar speaker appears more likely to activate an effect of behavioral mimicry than an agent speaker, probably because the avatar speaker provides richer nonverbal cues than the agent speaker. Detailed analysis of the motion data for speaker and listener head movements reveals systematic differences in a) head nodding between a speaker producing a monologue and a speaker engaged in a dialogue b) head nodding of speakers and listeners in the high and low frequency domain and c) the reciprocal dynamics of head-nodding with different virtual speaker's head nodding behavior. We conclude that: i) the activation of behavioral mimicry requires a certain number of nonverbal cues, ii) speakers behave differently in monologue and dialogue, iii) speakers and listeners nod asymmetrically in different frequency domains, iv) the coordination of head nods in natural dialogue is no more than we would expect by chance, v) speakers' and listeners' head nods become coordinated by spontaneous collaborative adjustment of their head nods.

Acknowledgements

The work is supported by EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1). Thanks to the Great Britain-China Educational Trust (GBCET) for their financial support in the last few months.

I want to thank my primary supervisor, Professor Patrick G.T. Healey, for his patient, kindness, guidance and complete support. I want to give great honor to my second supervisor Professor Peter William McOwan (1962-2019). He had been accommodating in every progress stage. I want to thank my external supervisor Dr. Marco Gillis and his wife, Dr. Xueni Pan. Thanks for their excellent help with my Ph.D. and research career. I want to thank my examiners Prof. Antonia Hamilton and Dr. Qianni Zhang for their elaborative review of the thesis.

I want to thank the administration team in the CDT Media and Arts Technology for their support, including Professor Nick Bryan-Kinns, Jonathan Winfield, Geetha Bommireddy, and Karen Bray.

I want to thank everyone in the **CogSci** at QMUL. Many thanks to Sophie Skach, Tom Gurion, Dr. Lida Theodorou, Soomi Park, Dr. Julian Hough, etc. for the great social life in the Cognitive Science research group. Thank all the MAT PhDs, including Dr. Yongmeng Wu, An Liang, Liang Men, Giulio Moro, Lucia Marengo, Thomas Deacon, Dr. Daniel Gabana, etc. We were progressing together. Thank all my Chinese friends in the UK. It was great to be friends with you and enjoy Chinese meals together. Thanks to all my senior friends, including Prof. Hao Tan, Prof. Zhengyu Tan, Dr. Ahn Sunghee, Mr. Park Soonpoong, for their help.

Last but not least, my gratitude and special thanks to my parents, Xizheng Zhang and Chunhua Liu, and my girlfriend Sisi Guo for their endless and unconditional love, accompany, patience, and encouragement along my way.

Contents

1 Introduction **15**

1.1 Motivation 15

1.1.1 Nonverbal Communication 15

1.1.2 Interpersonal Coordination 16

1.1.3 Persuasion 16

1.1.4 Immersive Virtual Environments 17

1.1.5 Digital Chameleons 18

1.1.6 Summary 18

1.2 Aims 19

1.2.1 Research Question 19

1.2.2 Research Goals 19

1.3 Method 20

1.4 Contributions 21

1.5 Publications 22

1.6 Thesis structure 22

2 Background **24**

2.1 Nonverbal Communication and Coordination 25

2.1.1 Nonverbal Communication 25

2.1.2 Head Nods 26

2.1.3 Speaker and Listener in Dialogue 27

2.1.4 Interpersonal Coordination 29

2.1.5 Summary 31

2.2 Behavioral Mimicry 32

2.2.1 Timing of Behavioral Mimicry 34

2.2.2 Chameleon Effect 34

2.2.3 Perception-Behavior Link 37

2.2.4 Summary 41

2.3 Behavior Research Method 41

2.3.1 Overview 41
2.3.2 Ethnomethodology 42
2.3.3 Confederate 43
2.3.4 Intelligent Virtual Agent 45
2.3.5 Immersive Virtual Environment 47
2.3.6 Summary 51
2.4 Digital Chameleons 52
2.5 Summary 54

3 Methodology 56
3.1 Replications of “Digital Chameleons” 57
3.2 Evaluating the Effectiveness of Virtual Speaker 58
3.2.1 Measures for Effectiveness 59
3.2.2 Principle Component Analysis 60
3.3 Head Nods Analysis 60
3.3.1 Descriptive Statistics of Head Nods 60
3.3.2 Head Nods by Frequency 61
3.3.3 Head Nods Coordination 62
3.4 Summary 71

4 Experimental Immersive Virtual Environment Design 73
4.1 System Architecture 74
4.2 Apparatus Settings 75
4.3 Virtual Environment 77
4.3.1 Virtual Character Library 79
4.3.2 Animating Virtual Characters 79
4.4 Motion Data Recording 81
4.5 Manipulating Head Movement 82
4.6 Multithreading 83
4.7 Evaluating the System 83
4.7.1 Supporting IVA Speaker-Listener Interaction . . . 83
4.7.2 Supporting Avatar Speaker-Listener Interaction . 88
4.8 Summary 91

5 Evaluating the “Digital Chameleons” Paradigm 92
5.1 IVA Speaker and Listener Interaction 93
5.1.1 Participants 93
5.1.2 Procedure 93

5.1.3 Effectiveness of the IVA Speaker 96
5.1.4 Results 96
5.1.5 Discussion 103
5.1.6 Conclusion 106
5.2 Avatar Speaker and Listener Interaction 106
5.2.1 Participants 107
5.2.2 Procedure 107
5.2.3 Pairing Participants 109
5.2.4 Effectiveness of the Avatar Speaker 110
5.2.5 Results 111
5.2.6 Discussion 115
5.2.7 Conclusion 116
5.3 Summary 116

6 Evaluating the Coordination of Head Nods 117
6.1 Head Nods in the IVA Speaker and Listener Interaction . 118
6.1.1 Descriptive Statistics of Head Nods 118
6.1.2 Head Nods Coordination 124
6.1.3 Discussion 132
6.1.4 Conclusion 133
6.2 Head Nods in Virtual Avatar Speaker-Listener Interaction 134
6.2.1 Descriptive Statistics of Head Nods 134
6.2.2 High and Low Frequency Head Nods 135
6.2.3 Head Nods Coordination 140
6.2.4 Discussion 146
6.2.5 Conclusion 148
6.3 Summary 148

7 Discussion 150
7.1 Summary of Hypotheses and Outcomes 151
7.2 Limitations of the "Digital Chameleons" Paradigm . . . 151
7.2.1 False Condition Design 152
7.2.2 Agents versus Avatars 155
7.2.3 Communicative Functions of Head Nods 155
7.2.4 Bidirectional Mimicry 156
7.2.5 Conclusion 157
7.3 Coordination of Nods in Dialogue 157
7.3.1 Collaborated Head Nods 157

7.3.2 Unnoticeably Fast Head Nods 158
7.3.3 Asymmetries in Head Nods between Speakers and Listeners 159
7.3.4 Coordinated Head Nods 160
7.3.5 Conclusion 162
7.4 Methodological Approach 162
7.4.1 Using Immersive Virtual Environment 163
7.4.2 Replicating the "Digital Chameleons" Paradigm . 165
7.4.3 Evaluating the Effectiveness of Virtual Speakers . 166
7.4.4 Behavior Analysis Methods 166
7.4.5 Conclusion 168
7.5 Summary 169

8 Conclusions and Further Work 170
8.1 Major Findings 171
8.2 Limitations and Future Works 172
8.3 Closing Remarks 175

Bibliography 176

Appendices 204

A Training message 205

B The Persuasive Message of Student ID Card 206

C Questionnaires 208
C.1 Demography 208
C.2 Agreement 208
C.3 Social Presence 209
C.4 Impressive 210
C.5 Realism 210
C.6 Overall 210

D Information Sheet and Consent Form A 213

E Information Sheet and Consent Form B 216

List of Figures

2.1 Number of times participants mimicry behaviors (rub their face or shake their foot) after the confederates' exact behavior. 36
2.2 The experimental set-up for the settings of the 'Digital Chameleons'. 53

3.1 Trajectory of Lorenz Attractor in 3 Dimension Space . . 66
3.2 Trajectory of Head Movement in 3 Dimension Space . . . 66
3.3 Reconstruct Lorenz Attractor in the 3 Dimension Phase Space 67
3.4 Reconstruct Head Movement in the 3 Dimension Phase Space 68
3.5 Creating Recurrent Plot with Thresholding Radius . . . 69
3.6 The Recurrent Plot of Head Nodding 70
3.7 The Colored Recurrent Plot of Head Nodding 70

4.1 Physical Architecture of the System 74
4.2 Schematic Diagram of the Capture Set-up in the Human Interaction Lab 75
4.3 Participant wore black suits attached with markers and Oculus Rift DK2 (HMD) 76
4.4 The virtual environment in the participant's view 78
4.5 Use Adobe Fuse to Create 3D Characters 80
4.6 Captured Body Movement as Modeled in Different Stages of Re-targeting and Re-synthesis 80
4.7 Pipeline of Head movement Manipulations 83
4.8 Head Movement of Participant in One Minute 84

5.1 A Participant Is Exploring the IVE and Motion Tracking 94
5.2 The IVA Is Delivering the Message to A Participant . . . 94

5.3 The Boxplots of the Components of the Agent's Effectiveness Measures 97
5.4 The Error-Bar Chart for Authority in Three Conditions. 104
5.5 Two Participants Are Doing the Experiment 108
5.6 The Real Listener, Speaker and the Virtual Avatar of the Speaker 109
5.7 The Procedure Flow of Experiment II 110
5.8 The Boxplots of the Components of the Speaker's Effectiveness Measures 112
5.9 The Error-Bar Chart for the Factor of Effectiveness in the Four Conditions. 114

6.1 Histograph of Maximum Head Nodding 119
6.2 Gender Difference in Maximum Head Nodding 121
6.3 The Boxplot of Participants' Log Total Head Nods. . . . 125
6.4 Examples of Fluctuation Plot of Delay 126
6.5 The Plot of Embedding Dimension on %FNN 127
6.6 The Colored Recurrent Plots in the Mimic, Playback and Recording Conditions 128
6.7 The SD and Total Head Nods for Listener and Speaker . 136
6.8 Boxplots of the Cumulative Amount of Head Nods for the Virtual and Real Speaker-Listener Pair 138
6.9 Cumulative Mean Difference of the Amount of Head Nods for the Virtual Listener-Speaker Pair 139
6.10 Cumulative Mean Difference of the Amount of Head Nods for the Real Listener-Speaker Pair 140
6.11 The Colored Recurrent Plots for Virtual Speaker-Listener Pair 142
6.12 The Colored Recurrent Plots for Real Speaker-Listener Pair 143
6.13 Boxplots of xRQA Outputs for the Virtual and Real Speaker-Listener Pair 144

7.1 The Mean Size (Degree) for Slow and Fast Nods 159

List of Tables

4.1 The Equipment Used in the Study 75
4.2 Two Data Log Methods Comparison 81
4.3 The Effectiveness of Speaker in Natural Condition. . . . 89
4.4 The Answers for the Open Questions. The number in the bracket represents how many times different participants repeated the answer. 90

5.1 Number of Participants in Each Condition by Experiment Trial 95
5.2 The Components of the Agent's Effectiveness Measures . 96
5.3 Model Comparison of GLMM for the Agent's Effectiveness Measures 98
5.4 Test of Fixed Effects for the Agent's Effectiveness Measures 98
5.5 Estimates of Fixed Coefficients for the Agent's Social Presence 98
5.6 Estimates of Covariance for the Agent's Social Presence . 98
5.7 Pairwise Comparison of Agent's Social Presence by Condition 99
5.8 Pairwise Comparison of Agent's Social Presence by Order 99
5.9 Factor Components and Loadings in IVA-Listener Interaction. 100
5.10 The Descriptive Statistics of The Main Factors. 101
5.11 Model Comparison of GLMM for the Main Factors . . . 101
5.12 Fixed Effects in GLMM for the Main Factors 102
5.13 Estimates of Fixed Coefficients for Authority 102
5.14 Estimates of Covariance for Authority 102
5.15 Pairwise Comparison of Agent's Authority by Condition 103
5.16 The Manipulation of the Virtual Speaker by Condition . 109
5.17 Number of Participants in Each Condition 109
5.18 Results for the Components of the Speaker's Effectiveness Measures 111

5.19 Factor Components and Loadings in Speaker-Listener Interaction. . . . 113
5.20 The Result of One-way ANOVA on The Factors between Conditions. . . . 114

6.1 The Descriptive Statistics for Participant's Maximum Head Nodding in Different Agent's Behavior Condition 119
6.2 Model Comparison of GLMM for the Logarithmic Maximum Head Nodding 120
6.3 Test of Fixed Effects for the Logarithmic Maximum Head Nodding 120
6.4 Estimates of Fixed Coefficients for the Logarithmic Maximum Head Nodding 120
6.5 Estimates of Covariance for the Logarithmic Maximum Head Nodding 121
6.6 Pairwise Comparison of the Logarithmic Maximum Head Nodding by the Participant's Gender 122
6.7 The Descriptive Statistics for Participant's Logarithmic Total Head Nodding in Different Agent's Behavior Condition 122
6.8 Model Comparison of GLMM for the Logarithmic Total Head Nodding 122
6.9 Test of Fixed Effects for the Logarithmic Total Head Nodding 123
6.10 Estimates of Fixed Coefficients for the Logarithmic Total Head Nodding 123
6.11 Estimates of Covariance for the Logarithmic Total Head Nodding 123
6.12 Pairwise Comparison of the Logarithmic Total Head Nodding by the Participant's Order 124
6.13 The Descriptive Statistics for %REC, %DET and LMAX in Different Agent's Behavior Condition 129
6.14 Model Comparison of GLMM for the xRQA Outputs 130
6.15 Test of Fixed Effects for the xRQA Outputs. The mimic condition is excluded for the test of LMAX because the LMAX in mimic condition is extremely (100x) higher than the other two conditions. 131
6.16 Pairwise Comparison of the %REC 131
6.17 Pairwise Comparison of the %DET 131

6.18 Pairwise Comparison of the LMAX 132
6.19 The ANOVA tests for Listeners' SD and Total Head Nodding in Different Agent's Behavior Condition 135
6.20 Paired t-tests: The Difference between Speaker and Listener's Head Nods in Different Condition 137
6.21 ANOVA Tests for xRQA Outputs of Virtual Speaker-Listener Pair 143
6.22 Cross Recurrence Quantification Analysis of Real Speaker-Listener Pair 145

7.1 The Hypotheses About the Relationship Between Mimicry and Effectiveness in the Thesis and Their Outcomes . . . 151
7.2 The Hypotheses About Head Nodding in the Dialogue in the Thesis and Their Outcomes 152
7.3 The Hypotheses About the Coordination of Head Nods in the Dialogue in the Thesis and Their Outcomes 153

List of abbreviations

IVE	Immersive Virtual Environment
IVA	Intelligent Virtual Agent
VR	Virtual Reality
DOF	Degree of Freedom
ROM	Range of Movement
HMD	Head Mounted Display
TSI	Transformed Social Interaction
RQA	Recurrence Quantitative Analysis
xRQA	Cross Recurrence Quantitative Analysis
RP	Recurrence Plot
DM	Distance Matrix
RM	Recurrence Matrix
SD	Standard Deviation
SE	Standard Error
AMI	Average Mutual Information
FNN	False Nearest Neighbor

Chapter 1

Introduction

1.1 Motivation

1.1.1 Nonverbal Communication

Human communicate with each other in verbal and nonverbal ways. While verbal communication refers to spoken language in tone and vocal, nonverbal communication can demonstrate a much broader range. It consists of the interaction with head movements, hand gestures, facial expression, interpersonal distance, gaze, postures, and voice intonation. These behaviors are often subconscious and unintentional (Zajonc, 1980). Many studies have proved that nonverbal behaviors serve as an essential part of the human interaction (Hinde, 1972; Kendon, 1990; Scheflen, 1964).

Head movements, among all these nonverbal behaviors, are more restricted. It has only 3 degrees of freedom (DOF), which are head pitch, yaw, and roll (or 6 DOF considering the limited mobility in the three directions on top of the neck). It has been shown, and is cultural convention, that people use head nods for agreement and head shakes for disagreement. Nevertheless, head movements can communicate more than that. For example, for speakers, head nods can further signal emphasis, the rhythm of speech; for listeners, head nods can indicate the level of understanding, floor taking (Hadar, Steiner, & Rose, 1985). The effectiveness of communication via head movements largely depends on the timing or the coordination between speaker and listener in the dialogue.

1.1.2 Interpersonal Coordination

Interpersonal coordination (Bernieri & Rosenthal, 1991) is usually referring to interactional synchrony and behavioral mimicry. It can sometimes be confusing to distinguish these terms in detail. Both of them indicate that people doing the same or similar actions as the other people, at roughly or precisely the same time. Behavioral mimicry is typically assessed by looking to the behaviors that are repeated in a short window (around 2 to 10 seconds) by other people. On the other hand, interactional synchrony mainly focuses on the timing or rhythm of the coordinated movements of the interactants (Lakin, 2013).

Interactional synchrony and behavioral mimicry can be subconscious and unintentional. Chartrand and Barge (1999) suggested that people automatically mimic each other's movements and behaviors unconsciously during an interaction, usually within a short window of time between three to five seconds. They claimed that automatic behavioral mimicry could: 1) prompt changes in individuals' cognitive processing style, 2) altering performance on tests of ability and creativity, 3) shifting preferences for consumer products, as well as 4) improving liking, empathy, affiliation, 5) increasing help behavior, and 6) reducing prejudice between conversational partners (Chartrand & Lakin, 2013). Based on this idea, Bailenson and Yee conducted the "Digital Chameleons" study (Bailenson & Yee, 2005). They created an intelligent virtual agent (IVA) speaker that was automatically mimicking the listener's head nods. Having compared the effectiveness of the mimicking IVA with the non-mimicking IVA, they suggested that the mimicking IVA is more persuasive than the non-mimicking IVA when presenting the regulation of student ID card.

1.1.3 Persuasion

Persuasion has naturally attracted attention in studies of nonverbal behavior because of the potential practical implications for human interaction in business, politics, and everyday life. Evidence from social science suggests that, overall, persuasive communicators tend to be more active with nonverbal behavior and more responsive to addressees (Mast & Cousin, 2013; Mehrabian, 2017). This positive change includes a range of nonverbal cues, including increased eye-contact, greater facial expressiveness, more frequent nods, and increased gesturing (Burgoon, Birk, & Pfau, 1990; Edinger & Patterson, 1983; Mast & Cousin, 2013; Mehrabian & Williams, 1969).

The studies that provide evidence for these conclusions face some critical methodological challenges. Nonverbal interaction is, of course, complex and dynamic. In natural communication, a range of different behaviors – facial expressions, gestures, nods, body position, body orientation – are concurrently deployed. It is hard to capture this complexity through detailed observation alone, and there is always some uncertainty about how well such observations generalize beyond the cases described (Kendon, 1990; Scheflen, 1964). One response to this is to generate hypotheses about specific nonverbal behaviors and investigate whether their frequency varies systematically with the rated persuasiveness of human speakers' performances (Burgoon et al., 1990; Mehrabian & Williams, 1969). This method can reveal systematic patterns but leaves questions of causation open.

Early experimental studies that attempted direct tests of the effects of nonverbal behaviors on persuasion typically tried to manipulate the speaker's performance. For example, giving instructions to the speaker to be more or less persuasive, or instructions about how to 'perform' effectiveness, such as using specific body angle, % gaze, and levels of gesturing and smiling (Edinger & Patterson, 1983). A known problem with this general approach is that 'actors' and confederates find it challenging to control specific aspects of their behavior in the live interaction. They could not only perform specific actions without influencing their performance of other tasks and undermining the spontaneousness of the communication in general (Bavelas & Healing, 2013; Kuhlen & Brennan, 2013). This problem is substantial, and some effects appear to be entirely the result of a confound due to the use of confederates (Bavelas & Healing, 2013).

1.1.4 Immersive Virtual Environments

Immersive Virtual Environments (IVEs) offer a potentially powerful way to address these difficulties because they provide experimenters with greater experimental control of individual behaviors (Bailenson, Blascovich, Beall, & Loomis, 2001). The step of rendering or animating actions into a virtual environment means that it is possible, in principle, to alter unusual movement selectively while leaving others unchanged. Based on this idea, Transformed Social Interaction (TSI) (Bailenson, 2006) was developed to investigate nonverbal interaction. TSI can have a significant impact on people's persuasive and instructional abilities by systematically filtering the appearance and behavior of their avatar in the eyes of their conversational partners, strategically amplifying or suppressing nonverbal signals in real-time.

1.1.5 Digital Chameleons

The "Digital Chameleons" studies are one early study of TSI studies in which an agent covertly mimics a human participant's head movements, at a short delay. At the same time, other aspects of message delivery remain constant (Bailenson & Yee, 2005). This study was designed as a test of Chartrand and Bargh's (1999) hypothesis. It suggests that automatic mimicry tends to increase affiliation between people. Automatic mimicry should, therefore, make an avatar more persuasive (Bailenson & Yee, 2005) and more likely to be trusted (Verberne, Ham, Ponnada, & Midden, 2013).

However, these results have not been consistently replicated (Hale & Hamilton, 2016; Riek, Paul, & Robinson, 2010). This shadows that the effect of behavioral mimicry is subtle or fragile than it is generally assumed. These studies might have ignored some critical aspects. For example, the speaker-listener interaction in the "Digital Chameleons" studies might not be realistic enough to recreate the natural mimicry behavior. Head nods are the central experimental manipulation of the "Digital Chameleons" studies. However, the coordination of head nods between speaker and listener in the natural interaction has not been considered in the experiments.

1.1.6 Summary

In summary, IVEs are becoming increasingly dominant in the research of nonverbal communication. IVEs enabled multimodal real-time full-body interaction and provided high experimental controllability in behavior research. However, the design of the virtual experiments should be carefully considered from the perspective of human interaction – or speaker-listener interaction in dialogue. Otherwise, like the "Digital Chameleons" studies, these experiments cannot avoid the replication crisis.

This thesis investigates the coordination of head nods in dialogue by recreating the experiments of the "Digital Chameleons" study.

1.2 Aims

The following section presents the overall research question of this thesis. Based on this research question, some more focused research goals are defined in detail.

1.2.1 Research Question

The overall research question this thesis addresses is: *How do speakers' and listeners' head nods become coordinated in dialogue?*

This paragraph specifies the meaning of the terminologies used, some more detail of their definition and origins are discussed in Chapter 2. The term *speaker* refers to the person who is delivering a message to the other people with the verbal and nonverbal signals. The term *listener* is the person who is receiving the message from the speaker; he or she could provide verbal or nonverbal feedback to the speaker. The verbal signal from the listener should be significantly less than the speaker. The term *head nods* refer to the head movements (rotation) in the direction of up and down. It is also named as head pitches. A head nod consists of a head rotates down first, followed with a head rotates up. The range of movement (ROM) should not exceed a certain degree, e.g., 180 degrees. The term *coordinated* refers to the behavioral matching/synchrony status of multiple subjects. The term *dialogue* refers to the communication process within just two people. One as the speaker, the other as the listener. The speaker and listener can switch roles at any time of the dialogue process.

1.2.2 Research Goals

Three specific research goals are unpacked concerning the overall research question.

1. **Developing a customized IVE for the replication of the "Digital Chameleons" study.**

Modern behavior research should take advantage of IVEs to enable multimodal real-time full-body interaction and controllability of experiments. As elaborated in the background (Section 2.3.4 and Section 2.3.5), the IVE should be integrated with a fine-grained full-body motion capture system and provide an interactive interface through/with an avatar or intelligent virtual agent. Specifically, for the replication of the "Digital Chameleons" study, the IVE should provide the ability to manipulate avatar/agent's behavior – head nods in this study.

2. **Evaluating the "Digital Chameleons" paradigm.**

   Since several studies relating to the "Digital Chameleons" failed to replicate the results of the original, it is worth to examine this paradigm again. These studies used an intelligent virtual agent (Section 2.3.4) as the speaker, but it does not address the interaction between speaker and listener in dialogue (Section 2.1.4). Thus, a live version of this paradigm, which exploited the synergy between avatars (Section 2.3.5.2), also needs to be evaluated.

3. **Examining the head nods coordination in the dialogue.**

   The "Digital Chameleons" studies manipulated the speaker's head nods without addressing head nods coordination between speaker and listener in natural dialogue. Head nodding is highly role-specific nonverbal behavior (Section 2.1.2). The inappropriate manipulation of head nods could lead to misunderstandings or seemingly unnatural / unconventional cues. Understanding the coordination of head nods of speaker and listener in the natural dialogue would be essential to the success of increasing the effectiveness of a mimicking speaker.

1.3 Method

Using a custom-built immersive virtual environment (IVE), the head-movement manipulation reported in the "Digital Chameleons" study is replicated (Bailenson & Yee, 2005). The environment uses virtual characters (agents/avatars) controlled by full-body motion capture and selective manipulations of movement is enabled by a server. The primary manipulation in the "Digital Chameleons" study is head movement. The usability of the IVE is evaluated with qualitative analysis on the the participants' opinion of the interaction in the IVE collected with the open questions in the online questionnaires., The effectiveness of virtual speakers with different head-nodding behaviors is tested with quantitative analysis on the participants' ratings of the virtual speaker's effectiveness collected with the online questionnaires.

Furthermore, participants' motion data is collected with the motion capture data. The head nods coordination is analyzed with the nonlinear method – Cross Recurrence Quantification Analysis on the motion data. The measures and statistic analysis methods are presented in Chapter 3.

1.4 Contributions

The thesis mainly contributes to two research fields:

First, it contributes to the field of human interaction with a systematic understanding of the coordination of head nods in dialogue and the methods for analyzing nonverbal behavior. Second, it contributes to the domain of virtual reality with the extensive application illustrated in this thesis for human behavior study. The primary contributions of this thesis are:

- A customized IVE architecture integrated with a full-body motion capture system that could be used to investigate human interaction by manipulating human behavior with a customizable server pipeline. The system can support real-time natural social interaction through virtual avatars with full-body tracking.

- An extensive exploration of the "Digital Chameleons" effect – where and why does automatic mimicry work and where not? The questions are answered with two replications of the original study with IVE. Virtual characters with full-body motion tracking, etc. is added to the original setup. The results suggest that the realisticness of nonverbal cues play an essential role in the activation of the prosocial effect of automatic behavioral mimicry.

- A statistical model of speaker and listener's head nods coordination in dialogue. This model demonstrates asymmetries between speaker and listener head nods in the high and low frequency domain in conversation. It also shows differences in the coordination of nods in natural interaction and in mimicked interaction. This model is a novel contribution to the study of human head nods.

- An exploration of statistical methods for the analysis of human behavior time series data. Traditional linear methods and novel nonlinear methods are compared. This thesis suggests the potentials of the nonlinear techniques such as Cross Recurrence Quantification Analysis (xRQA) and frequency analysis for the study of behavior time series data.

1.5 Publications

Leshao Zhang and Patrick G.T. Healey. 2018. Human, Chameleon or Nodding Dog? Virtual Experiments with Non-Verbal Persuasion. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18). ACM, New York, NY, USA, 428-436. DOI: https://doi.org/10.1145/3242969.3242998

Leshao Zhang and Patrick G.T. Healey. 2019. Coordination of Head Nods: Asymmetries between Speakers and Listeners. In Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue. SemDial, London, UK. URL: http://semdial.org/anthology/Z19-Zhang_semdial_0017.pdf

1.6 Thesis structure

The thesis is structured as follow:

Chapter 2 provides a comprehensive review of research in 1) nonverbal communication, including nonverbal behaviors, head nods and speaker-listener coordination, 2) behavioral mimicry, including the chameleon effect and the perception-behavior link, 3) behavior research methods, including the use of ethnomethodology, confederate, intelligent virtual agent and immersive virtual environment and, 4) "Digital Chameleons" studies.

Chapter 3 describes the method of this research. In this chapter, the method of the replications of the "Digital Chameleons" study is illustrated, the measurement of the effectiveness of the agent/avatar speakers, and finally, the statistic methods for head nods behavior analysis. The hypotheses under these measures are specified accordingly.

Chapter 4 describes the design of the IVE system for the study, including the architecture of the IVE system, the lab setting, the techniques used in the study, as well as two pilot studies for the evaluations of the system.

Chapter 5 describes the results of the two studies , including an experiment on the effectiveness of the agent speaker and an experiment on the effectiveness of the avatar speaker. The results from the two studies are compared.

Chapter 6 describes the results from the head nods analysis. In this chapter, we analyze the speaker and listener's head nods with the linear statistic method and the coordination of head nods with nonlinear analysis methods.

Chapter 7 discusses the findings in this study, including the limitation of the "Digital Chameleons" paradigm, the coordination of head nods in dialogue and the methodological approach of this study.

Chapter 8 concludes with the findings and the limitations of this work and suggests the implications for future work.

Chapter 2

Background

In this chapter, the literature relevant to this study including the theories and methods associated with the “Digital Chameleons” hypothesis is reviewed. In addition to this, some literature from the perspective of speaker-listener interaction is further reviewed. Specifically, the topic of nonverbal communication (Section 2.1.1), head nods (Section 2.1.2), speaker and listener in dialogue (Section 2.1.3) and interpersonal coordination (Section 2.1.4) is introduced first. Behavioral mimicry is covered next, including an introduction to behavioral mimicry (Section 2.2), the Chameleon Effect (Section 2.2.2) – from which the “Digital Chameleons” study borrowed the automatic mimicry idea, and the perception-behavior link (Section 2.2.3). Then, a few research methods used in the behavior research, including ethnomethodology (Section 2.3.2), confederate (Section 2.3.3), intelligent virtual agent (Section 2.3.4) and immersive virtual environment (Section 2.3.5) is reviewed. At last, the “Digital Chameleons” studies (Section 2.4) is detail investigated, and the inconsistent results in the related studies is discussed.

2.1 Nonverbal Communication and Coordination

2.1.1 Nonverbal Communication

Nonverbal communication is communication with body movements or non-linguistic behaviors other than words (Mehrabian, 2017). Humans have been using nonverbal behaviors to communicate with each other since they were still apes. It is much later that humans developed the capability to communicate with each other with language. Nonverbal communication is established earlier than verbal communication. This fact makes nonverbal communication more instinctual and involuntary than verbal communication.

It is claimed that the involuntary basis of nonverbal communication can be used to tell a person's real state of mind (Navarro, 2008). For example, in an interview of a rape case, the suspect denied involvement, saying that he had turned left and gone home while his hand gestured to his right. Eventually, it turned out that he was lying. Nonverbal behavior is believed to be more credible than verbal language. However, this does not mean that we can read other people's minds through their nonverbal behaviors since nonverbal behaviors are very ambiguous, on the other hand.

People communicate nonverbally through many ways including: **head movements**, e.g., nods, shakes, orientation (Pease & Pease, 2008), **facial expressions**, e.g., happiness, sadness, fear, anger, disgust (Andersen, 1999; Evans, 2002), **eyes**, e.g., contacts, vibration, direction (Andersen, 1999; Guerrero & Floyd, 2006; Martin & Nakayama, 2013), **voice**, e.g., pitch, volume, speaking rate (Andersen, 1999; Buller & Aune, 1988), **smell**, e.g., pheromones, perfume (Hickson, Stacks, & Moore, 2004; Thornhill & Gangestad, 1999; Wyatt et al., 2003), **gesture**, e.g., adaptors, emblems, and illustrators (Andersen, 1999; Krauss, Chen, & Chawla, 1996), **posture**, e.g., standing, sitting, squatting, and lying down (Hargie, 2016), **touch** (Andersen, 1999; Heslin & Alper, 1983; Jones, 1999), **proxemics**, e.g., public space, social space, personal space, intimacy space (Andersen, 1999; Hall et al., 1968), **time**, e.g., biological, personal, physical, and cultural time (Andersen, 1999), **personal presentation and environment** (Schmitz, 2012).

Nonverbal communication is developed with evolution, body structure, and social learning (Mehrabian, 2017). Nonverbal behavior can function 1) as the replacement verbal communication when impossible or inappropriate 2) as complementation to enhance verbal communication 3) as modification to the speech 4) as contradiction, either intentionally or unintentionally, to what is said 5) as regulation for speech turns 6) to express emotions and interpersonal attitudes 7) to negotiate relationships in respect of, e.g., dominance, control and liking 8) to convey personal and social identity through, e.g., dress and adornments 9) to contextualize interaction with particular social setting (Hargie, 2016).

Specifically, head nods among all the nonverbal behaviors is the focus in this study. Also, head nods are the central experimental manipulation in the "Digital Chameleons" study.

2.1.2 Head Nods

Head nods are an especially crucial conversational signal (Battersby & Healey, 2010a; Boholm & Allwood, 2010; Hadar et al., 1985; Heylen, 2005, 2006). They are the most frequent head movement behavior among shakes and changes of angle/orientation. (Ishi, Ishiguro, & Hagita, 2014a; Włodarczak, Buschmeier, Malisz, Kopp, & Wagner, 2012). One possible reason for the mixed evidence on head-nodding coordination is the potential for different kinds of nod with different frequencies.

Hadar et al. (1983) distinguished 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 suggested that listeners mainly use ordinary head nods to signal 'YES', fast 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. (2019) defined 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 serve different functions for speakers and listeners. For speakers, head nods can 1) serve as a signal of the intention to continue speaking, 2) to seek or check agreement, 3) to express emphasis, 4) to control and organize the interaction, 5) as 'beat' gestures that accompany the rhythmic aspects of speech (Ishi, Ishiguro, & Hagita, 2014b), 6) to signal lexical repairs and 7) mark switches between direct and in-direct discourse. For listeners, head nods serve as 1) 'backchannels' to signal their level of understanding of an ongoing turn, 2) a signal that a (currently unaddressed) listener would like to take the floor, 3) to communicate the degree of understanding, agreement, or support (Hadar et al., 1985; Heylen, 2005). Listener head nods can provide concurrent 'backchannel' feedback to a speaker's turn (Yngve, 1970). Single head nods or jerks are the most frequent feedback movements. Minimal head nods are used to show the continuation of contact, perception, and understanding. More complex head movements are used for emphasis, agreement, self-reflection, citation, self-reinforcement, and own communication management (Boholm & Allwood, 2010), disagreement, surprise, and disappointment. Multiple nods or sequences of expressions, e.g., nods and smiles, can be used to acknowledge/refuse an idea or to ask for clarification (Allwood & Cerrato, 2003).

Head nods are also different in quantity for speakers and listeners. 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. However, Hale et al. (2019) suggested that listeners nod more in the high-frequency domain.

2.1.3 Speaker and Listener in Dialogue

Speakers and listeners behave in systematically different ways. For example, speakers gesture more and addresses gesture systematically less than unaddressed third parties (Battersby & Healey, 2010b; Gerwing & Bavelas, 2013; Healey et al., 2013; Healey, Plant, Howes, & Lavelle, 2015). On the other hand, speaker and listener's nonverbal behaviors are tightly coupled (Richardson, Dale, & Kirkham, 2007). Speaker and listener's verbal and nonverbal behaviors incrementally establish the conversation through their reciprocal dynamics. In other words, both the speaker and the listener's verbal and nonverbal behaviors shape the dialogue.

Then, what is a dialogue? Clark (1996) outlined of 10 essential features of face-to-face dialogue:

1. Co-presence: Both participants are in the same physical environment.

2. Visibility: They can see each other.

3. Audibility: They can hear each other.

4. Instantaneity: They see and hear each other with no perceptible delay.

5. Evanescence: The medium does not preserve its signals, which fade rapidly.

6. Recordlessness: Their actions leave no record or artifact.

7. Simultaneity: Both participants can produce and receive at once and simultaneously.

8. Extemporaneousness: They formulate and carry out their actions spontaneously, in real-time.

9. Self-determination: Each participant determines his or her actions (which means not scripted).

10. Self-expression: The participants engage in actions as themselves (which means not act as roles).

This definition of face-to-face dialogue is rather rigorous. Gerwing et al. (2013), on the other hand, described what is not a dialogue: a speaker who is alone in the lab, describing something to a camera with no addressee. They also emphasized the extemporaneousness, self-determination, and self-expression of the real dialogue.

One of the oldest models of dialogue is as a unidirectional transmission of information from a sender to receiver in the classic message model (Shannon, 1948). In this model, the sender will encode the message and deliver it to the passively listening receiver. The receiver will then decode the message and await a speaking turn. A listener is assumed to be relatively passive. Conversational partners are considered as either 'speaker' or 'speaker-in-waiting' (Bavelas, Coates, & Johnson, 2000). Dialogue is regarded as a sequence of alternating monologues in which speakers take turns talking. This naïve model greatly simplifies the communication process so that it fails to depict the pragmatic dialogue process.

Dialogue is considered fundamentally different from a monologue in the interactively aligned partner view. In this model, the presence of a conversational partner changes the cognitive processes for language production and comprehension. Conversation is a joint activity, and speaker and listener interaction is an interactive alignment process (Garrod & Pickering, 2004; Pickering & Garrod, 2004b). In this process, a speaker's utterances automatically activate the same ideas, words, or syntactic structures in the mind of an addressee through a fast, inflexible, unconscious priming process. Dialogue partners automatically mimic each other's communicative behaviors (verbal or nonverbal) (Lakin & Chartrand, 2003). Dialogue partners' mental representations are automatically aligned through the perception-behavior link (Chartrand & Dalton, 2009; Dijksterhuis & Bargh, 2001). This process greatly simplifies production and comprehension in dialogue (Pickering & Garrod, 2004b). However, it deemphasizes the social nature of dialogue and the moment-by-moment coordination among partners.

On the contrary, in the collaborative-partner view, listeners can also make significant contributions to speakers' utterances as co-creators or co-narrators (Bavelas et al., 2000; Krauss, 1987; Wilkes-Gibbs, 1986). In communication, listeners will show their needs, comprehension, and feedback in the grounding process with verbal and nonverbal behavior. For example, in a movie description task, with more feedback received from the listener, the speaker's narratives became more comprehensible, and the listener understood better as well (Kraut, Lewis, & Swezey, 1982). Listeners are not passive recipients. They can actively shape the interaction. Speakers regularly monitor their listeners for concurrent nonverbal feedback. When listeners' nonverbal feedback is absent, mistimed, or incongruent, e.g., raised eye-brows or puzzled looks when a smile was expected, speakers adjust their performance - mid-turn - to try and get things back on track (Bavelas et al., 2000; Goodwin, 1979). Authentic speakers should take into account their listener's informational needs and to fit their utterances accordingly (Bell, 1984; Clark & Murphy, 1982; Clark & Carlson, 1982; Fussell & Krauss, 1992; Lockridge & Brennan, 2002).

2.1.4 Interpersonal Coordination

Conversation is a joint activity that requires interpersonal coordination in the theory of dialog (Cappella, 2005; Clark, 1996). Interpersonal coordination suggests that behaviors are often patterned and synchronized in social interactions. People are doing the same or similar actions as the other people at roughly or precisely the same time (Chartrand &

Lakin, 2013). In the literature, numerous terms can be found related to interpersonal coordination, including chameleon effect, mimicry, social resonance, coordination, interactional synchrony, attunement. Usually, interpersonal coordination is considered to be interactional synchrony and behavioral mimicry (Bailenson, Yee, Patel, & Beall, 2008; Bernieri & Rosenthal, 1991; Chartrand & Lakin, 2013; Lakin, 2013). In this section, we mainly talk about the interactional synchrony. Behavioral mimicry is explained in detail in Section 2.2.

Interactional synchrony is often related to contingency, mutuality, mutual adaptation reciprocity, mutual responsiveness, affect attunement, dyadic synchrony, dyadic affect regulation, and behavioral entrainment. It is the reciprocal dynamic of the temporal structure of behaviors between interactants. It focuses on the timing and rhythms of the behaviors (Condon & Ogston, 1967) and the degree of congruence between the behavioral cycles of engagement and disengagement of two people. On the other hand, behavioral mimicry research is interested in the nature of the behaviors. In Bernieri's view (1991), interactional synchrony is "the degree to which the behaviors in an interaction are non-random, patterned or synchronized in both form and timing". Harrist & Waugh (2013) suggested that interactional synchrony emerges from 1) keeping mutual attention for a long time, and "tracking each other" 2) temporal coordination of behaviors such as body orientation, body movements, facial expressions 3) contingency and 4) attunement – synchrony in infant-adult caregiver interactions.

Interactional synchrony can be found in many phenomena, including **leg movements** (Schmidt & Carello, 1990; van Ulzen, Lamoth, Daffertshofer, Semin, & Beek, 2008), **body posture sway** (Shockley, Santana, & Fowler, 2003; Varlet, Marin, Lagarde, & Bardy, 2011), **eye movements** (Richardson & Dale, 2005), **hand clapping** (Néda, Regan, Bréchet, Vicsek, & Barabasi, 2000), **rocking chair movement** (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2008), **waving** (Lakens, 2010), **finger tapping** (Oullier, Guzman, Jantzen, Lagarde, & Kelso, 2008), **piano playing** (Keller, Knoblich, & Repp, 2007), **dancing** (Kirschner & Tomasello, 2010), etc. Interactional synchrony can be facilitated by interpersonal relationship. For example, mothers were more in sync with their own children (Bernieri, Reznick, & Rosenthal, 1988) and teachers were more in sync with their own students (Bernieri, 1988).

Interactional synchrony can increase affiliation (Hove & Risen, 2009), rapport (Mazzurega, Pavani, Paladino, & Schubert, 2011; Paladino, Mazzurega, Pavani, & Schubert, 2010; Vacharkulksemsuk & Fredrickson, 2012), prosocial behavior such as cooperation (Valdesolo, Ouyang, & DeSteno, 2010; Wiltermuth & Heath, 2009), helping behavior (Kirschner & Tomasello, 2010).

The distinction of the definition between interactional synchrony and behavioral mimicry is rather vague (Ramseyer & Tschacher, 2006). Both the causes and the consequences of interactional synchrony and behavioral mimicry are often similar (Chartrand & Lakin, 2013; Delaherche et al., 2012). This similarity is also emphasized in Section 2.2. Interactional synchrony and behavioral mimicry can often be observed simultaneously. For example, two seated people with crossed legs or same gaze direction are considered as behavioral mimicry. It becomes interactional synchrony if they cross or uncross their legs at the same time or shift their gaze in the same direction simultaneously. These similarities and simultaneities suggest that both interactional synchrony and behavioral mimicry reliably serve the goal of interpersonal coordination in a more broadly purpose – to facilitate and regulate the varied and complex social interactions (Chartrand & Van Baaren, 2009; Knoblich & Sebanz, 2006; Marsh, Richardson, & Schmidt, 2009).

2.1.5 Summary

In this section, the nonverbal communication is briefly introduced. Nonverbal communication has many behavior channels. It has been used since our ancestors have not yet developed our languages. Now it serves as a complementary function of verbal communication. Head nodding, as an important nonverbal behavior, seems to be very different between speaker and listener in dialogue. While speakers and listeners communicate in a coordinated way, how does head nods coordinated between speakers and listeners in the dialogue is not yet clear. Interpersonal coordination refers to interactional synchrony and behavioral mimicry. Interactional synchrony is explained in this section. In the next section, behavioral mimicry is investigated.

2.2 Behavioral Mimicry

Behavioral mimicry is the automatic imitation of gestures, postures, mannerisms, and other motor movements (Chartrand & Lakin, 2013). The word 'automatic' suggests that the behavioral mimicry is nonconscious, unintentional, and effortless. Mimicry serves as a communicative function that communicates understanding and togetherness, thus creates empathy and rapport between interaction partners that lead to positive social outcomes (Bavelas, Black, Lemery, & Mullett, 1986; Bernieri, 1988; LaFrance, 1979, 1982).

Mimicry has been claimed for a wide range of behaviors, including **yawning** (Helt, Eigsti, Snyder, & Fein, 2010; Provine, 1986), **body posture** (LaFrance, 1982; Tia et al., 2011; Tiedens & Fragale, 2003), **face touching** (Chartrand & Bargh, 1999; Lakin & Chartrand, 2003; Stel & Vonk, 2010; Yabar, Johnston, Miles, & Peace, 2006), **foot shaking** (Chartrand & Bargh, 1999; Lakin, Chartrand, & Arkin, 2008), **food consumption** (Herrmann, Rossberg, Huber, Landwehr, & Henkel, 2011; Johnston, 2002; Tanner, Ferraro, Chartrand, Bettman, & Baaren, 2007), **pen playing** (Stel & Vonk, 2010; van Baaren, Fockenberg, Holland, Janssen, & van Knippenberg, 2006), **coloring** (van Leeuwen, Veling, van Baaren, & Dijksterhuis, 2009), **handshake angle and speed** (Bailenson, Yee, Brave, Merget, & Koslow, 2007), **cospeech gestures** (Goldin-Meadow & Alibali, 2013; Holler & Wilkin, 2011b), **smoking** (Harakeh, Engels, Baaren, & Scholte, 2007), **eating** (Hermans et al., 2012), **finger tapping** (van Leeuwen, van Baaren, Martin, Dijksterhuis, & Bekkering, 2009), **facial expressions** (Bavelas et al., 1986; Dimberg, Thunberg, & Elmehed, 2000; Lundqvist & Dimberg, 1995), **emotional reactions of interaction partners** (Hatfield, Rapson, & Le, 2009; Hatfield, Cacioppo, & Rapson, 1993; Hawk, Fischer, & Van Kleef, 2011; Huntsinger, Lun, Sinclair, & Clore, 2009; Neumann & Strack, 2000), verbal characteristics of interaction partners, including **accents** (Giles, Coupland, & Coupland, 1991), **linguistic style** (Ireland & Pennebaker, 2010; Niederhoffer & Pennebaker, 2002), **speech rate** (Webb, 1969), and **syntax** (Levelt & Kelter, 1982).

A variety of effects have been claimed for behavioral mimicry, for example, **mimicry changes the cognitive processing** – people being mimicked becomes more context dependent (Van Baaren, Horgan, Chartrand, & Dijkmans, 2004), assimilative (Van Baaren, Janssen, Chartrand,

& Dijksterhuis, 2009), convergent thinking (Ashton-James & Chartrand, 2009), self-focus and self-consciousness (Gueguen, 2011); **mimicry increases persuasion and consumer behavior** – people agree more with the mimicker than non-mimicker (Bailenson & Yee, 2005; van Swol, 2003; Wong, Hartley, & Tombs, 2017); people being mimicked are more likely to buy a product (Herrmann et al., 2011; Jacob, Guéguen, Martin, & Boulbry, 2011; Stel, Mastop, & Strick, 2011; Tanner et al., 2007); **mimicry increases self-regulatory ability** – people gain better self-regulatory ability and self-control (Dalton, Chartrand, & Finkel, 2010), and fine-motor control (Finkel et al., 2006); **mimicry extends embodied cognition** – inappropriate mimicry lead to disliking (Stel et al., 2010), social coldness and physical chill (Bargh & Shalev, 2012; Leander, Chartrand, & Bargh, 2012; Zhong & Leonardelli, 2008); **mimicry increases liking and empathy** (Bavelas et al., 1986; Bernieri, 1988; Charny, 1969; Chartrand & Bargh, 1999; Holler & Wilkin, 2011b; Lafrance & Broadbent, 1976; Maurer & Tindall, 1983; Scheflen, 1964); **mimicry increases helping behavior** (Fischer-Lokou, Martin, Guéguen, & Lamy, 2011; Stel, Van Baaren, & Vonk, 2008; van Baaren, Holland, Kawakami, & van Knippenberg, 2004); **mimicry increases interdependence and feelings of closeness** – people being mimicked becomes more interdependent, self-construal (Redeker, Stel, & Mastop, 2011; Van Baaren, Maddux, Chartrand, De Bouter, & Van Knippenberg, 2003), and support for liberal groups(Stel & Harinck, 2011); **mimicry increases the accuracy in emotion perception** (Neal & Chartrand, 2011); **mimicry reduces prejudice** (Dalton et al., 2010; Inzlicht, Gutsell, & Legault, 2012).

Research also suggest that behavioral mimicry can be moderated by a lot of factors. Some of these factors act as facilitators and lead to more mimicry, including **pre-existing rapport** (Likowski, Mühlberger, Seibt, Pauli, & Weyers, 2008; McIntosh, 2006; Stel et al., 2010; Tickle-Degnen, 2006), **goal to affiliate** (Lakin & Chartrand, 2003; Over & Carpenter, 2009), better ability of **perspective taking** (Chartrand & Bargh, 1999) and **interdependent self-construal** (Van Baaren et al., 2003), **similarity** in opinions (Swol & Drury-Grogan, 2017) and knowledge (Castelli, Pavan, Ferrari, & Kashima, 2009; Clark & Kashima, 2007), positive **mood and emotion** (Likowski et al., 2011; van Baaren et al., 2006), **executive functioning** that cognitively occupied or under cognitive load (van Leeuwen, van Baaren, et al., 2009), **social intention** (Wong et al., 2017). The other factors act as inhibitors and lead to less mimicry, including **goal to disaffiliate** (Johnston, 2002; Yabar et al., 2006), negative **mood and emotion** (Likowski et al., 2011).

2.2.1 Timing of Behavioral Mimicry

Behavioral mimicry is a form of interpersonal coordination. Compared to interactional synchrony – the other type of interpersonal coordination, behavioral mimicry is commonly assessed by identifying that people are engaging in similar action or that an interaction partner repeats a particular behavior within three to five seconds delay (Chartrand & Lakin, 2013). Others suggested different delays for mimicry. For example, Baaren et al. (2009) and Leanders et al. (2012) suggested that mimicry occurs with 3-4 seconds delay; Stel et al. (2009) assessed mimicry if participants repeated an action in 10 seconds. Bailenson et al. (2004) checked mimicry with 1, 2, 4, and 8 seconds delay and suggested that mimicry with 1 second delay is easy to be detected. In the "Digital Chameleons" study, 4 seconds delay was used for automatic mimicry (Bailenson & Yee, 2005). Recently, Hale et al. (2016) checked the virtual mimicry's effect on rapport and trust using 1 or 3 seconds delay. They suggested that mimicry with 3 seconds delay has stronger effect on rapport than mimicry with 1 second delay. However, more studies did not specify the timing of mimicry at all. Overall, the timing of assessing behavioral mimicry in the previous studies varies from 1 second to 10 seconds. However, the selection of the timing of mimicry is arbitrary.

2.2.2 Chameleon Effect

The Chameleon Effect (1999) was first introduced by Chartrand & Bargh. It was used to describe nonconscious mimicry behavior – like a chameleon changing its color to match its current surroundings, humans altering their behavior to blend into social environments. This is in contrast to "monkey see, monkey do" or "to ape" which means to imitate intentionally.

Chartrand & Bargh claimed that the chameleon effect is the mechanism behind mimicry and behavioral coordination, and it causes interpersonal rapport and empathy. The perception-behavior link is claimed to be the reason for the chameleon effect. In the hypothesis of the perception-behavior link, the perception of another's behavior, e.g., facial expression, body posture, hand gesture, head nodding, increases the tendency for the perceiver to behave similarly, passively and nonconsciously. Moreover, perception causes similar behavior regardless of interpersonal goal or pre-existing interpersonal relationship, and the perception of the same behavior creates empathy and rapport.

Chartrand & Bargh did three experiments to test the phenomenon and effect of nonconscious mimicry and the individual difference with nonconscious mimicry. These experiments shared the same procedure but used different measures. In the experiments, they asked the participant and the confederate to describe the content in photos together. Confederates were instructed to vary their facial expression (smile or neutral) and their behavior (rub their face or shake their foot).

In the first experiment, they counted the number of times of participants' mimicry behavior, e.g., the number of times of the participants rubbed their face or shook their foot right after the confederates' exact act. The results showed that there was no significant effect for confederate facial expression on the counts of the behaviors being mimicked. However, participants rubbed their faces significantly more than chance when the confederates rubbed their faces. Participants shook their foot significantly more than chance when the confederates shook their foot (Figure 2.1). Chartrand & Bargh also reported that no participants noticed that the confederates were shaking their foot or rubbing their face. They suggested that 1) mimicry behaviors happens nonconsciously 2) nonconscious mimicries happens without affiliation goal (whenever the confederates smile or not) 3) nonconscious mimicries occurs among strangers (no pre-existing relationship needed) 4) nonconscious mimicries occurs at greater than chance level.

In the second experiment, they measured the likeability of the confederate and the smoothness of the interaction with the same task. The results showed that participants liked the confederates more when the participants were mimicked than they were not mimicked. The interactions between participants and confederates were reported smoother when the participants were mimicked by the confederates than they were not mimicked. These effects were not moderated by the behavior type. They also checked the confederates' behavior to ensure that no bias is be-

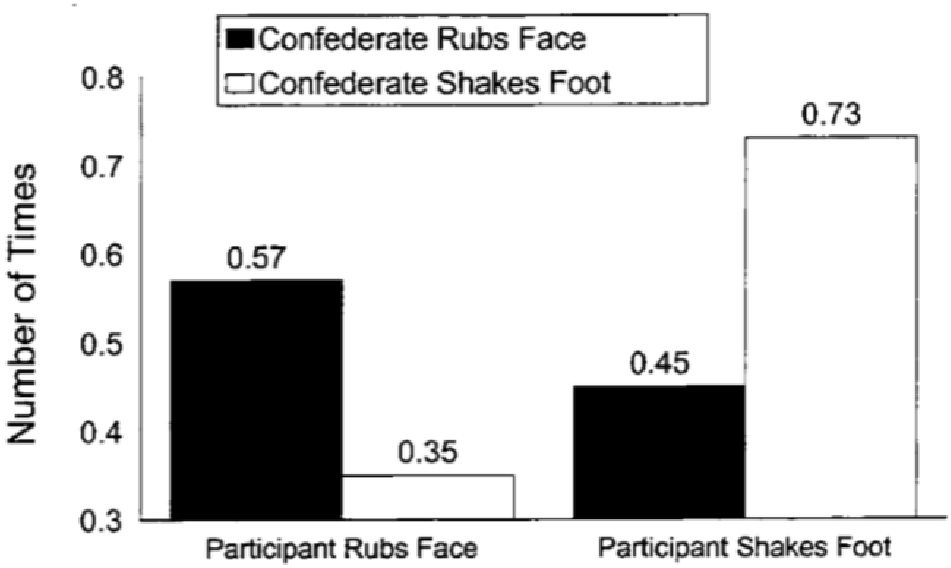


Figure 2.1: Number of times participants mimicry behaviors (rub their face or shake their foot) after the confederates' exact behavior. Adapted from Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910

ing introduced from the confederates. In previous research, mimicry has been suggested to be by-products or outgrowths of pre-existing rapport or liking (Levenson & Ruef, 1997; Scheflen, 1964); mimicry and rapport are positively correlated for people known to each other (LaFrance, 1979, 1982; Lafrance & Broadbent, 1976) and negatively correlated for people are strangers (Bernieri, 1988; La France & Ickes, 1981). However, with this study, Chartrand & Bargh claimed that the chameleon effect operates in a passive, goal independent manner, and creates greater liking and harmonious of interaction. They further suggested that nonconscious mimicry only happens in the context of social interaction and not happens between strangers with no interaction.

In the third experiment, they measured the perspective-taking ability and the empathic concern (an emotional concern-for-others facet of empathy) of the participants with the subscale of Davis's Interpersonal Reactivity Index (IRI) (Davis et al., 1980) as well as the number of times of participants' mimicry behavior. They defined the high perspective-taking for participants whose IRI subscale score above the median score, and low perspective taking for those below the median score. The same split was done with empathic concern. Their results showed that participants with high perspective-taking ability did more mimicry behaviors (rub their face or shake their foot) than those with low perspective-taking. However, no significant effect was found between empathic concern and mimicry behavior.

With the three experiments, Chartrand & Bargh suggested that 1) there exist the chameleon effect (nonconscious mimicry) 2) chameleon effect causes rapport and smoothness of the interaction between interactant 3) the perspective-taking ability of the individual moderates the chameleon effect. They further suggested that the perception-behavior link is the reason for chameleon effect.

2.2.3 Perception-Behavior Link

In Section 2.2.2, the perception-behavior link is the mechanism behind the chameleon effect is introduced. In this section, the perception-behavior link, and its theoretical basis is elaborated.

2.2.3.1 Direct Perception-Action Link

Perception is made for action. It is the guidance and control device of action. We all know that most plants do not have the perception system, and they do not move, while animals do. Ascidiacea, for example, will abandon its perception system when it finds a solid surface to attach to, and will not move anymore in the rest of its life. Without the need for action, perception is not needed anymore (Dijksterhuis & Bargh, 2001).

Perception can directly lead to action. The direct perception-action link is also called the long-term Stimuli-Response (S-R) bonds. It is derived from the prehistoric creatures that developed the nervous system to respond to the perceived stimuli from the environment. This direct link can also be learned over time based on one's history of reward and punishment stimuli (Gibson, 1978; McArthur & Baron, 1983; Skinner, 1938; Watson, 1913). That is, the direct perception-action link (long term S-R connection) can be either genetically prespecified or a product of learning and it is an automatic (unintentional) route.

On the other hand, there is short-term S-R connection. The short-term S-R connection is an intentional route of the response to a task-relevant stimulus. It is eatablished on the basis of task instructions and is held in short-term memory for the duration of the task (Barber &

O'Leary, 1997; Zorzi & Umiltá, 1995). Mimicry is an effect of the long-term S-R connection. It is a process modulated by attentional processes for input, and inhibitory processes for output (Heyes, 2011).

2.2.3.2 Human Inhibitory Processes

Unlike many other species (animals) of which perception always leads to action, human has the flexibility in that these action tendencies are not obligatory. Dijksterhuis (2001) explored the facilitator and inhibitor options and suggested that the human brain acting as the inhibitor, which leads to the flexibility of the action tendencies.

The facilitator option suggests that perception does not directly affect overt behavior unless an additional process, e.g., a consciously made decision or motivation, is performed. With the presence or absence of the facilitator, the perception will lead or not lead to action.

The inhibitor option, however, suggests that perceptual activity is sufficient to create action, but inhibitors can prevent this from occurring (Logan & Cowan, 1984). From the evolutionary perspective, the modern brain is developed by adding new parts to existing old ones (Dennett & Dennett, 1996; Milner & Goodale, 2006). So the old brain modules are still in our brain, i.e., direct perception-behavior links still exist. It can be moderated by new brain parts that have a certain degree of control over older ones. Furthermore, the studies of disordered people strongly favor the inhibitor option (Prinz, 1990). For example, frontal lobe damaged patients have less control over the direct effects of perception on behavior. They drink when they see water, eat when they see food even when it is inappropriate (Lhermitte, 1983; Passingham, 1993; Smith & Jonides, 1999). All this evidence suggests the inhibitor option is more likely to be the answer to the human flexibility to the direct perception-action link.

2.2.3.3 Overlapping Representations

Different from the direct perception-action link, the perception-behavior link is the perceiver's tendency to act in the same way as another's behavior. This tendency to imitate is suggested to be the result of the overlap of functioning area in our brains of the perceptual and behavioral representations, resulting in the automatic activation of the behavioral response by the perception of another's behavior, regardless that the behavioral response be stamped in as a habit through reinforcement and the response to be intended and strategic (Dijksterhuis & Bargh, 2001).

It is suggested that this tendency to imitate is a consequence of the shared representational systems for perception and action – the overlap of the mental representations between perception and behavior in primates as well as human beings (Chartrand, Maddux, & Lakin, 2005; Dijksterhuis & Bargh, 2001).

One century ago, Carpenter and James (1875; 1890) proposed the Ideomotor-action -– that merely thinking about doing something makes one more likely to act, i.e., "thinking is for doing". Later research provided evidence supporting this notion. For example, Paus et al. (1993) found that thinking about a word or a gesture activates the same area in the anterior cingulate cortex as actually uttering the word or making the gesture. Jeannerod (1994; 1997) claimed that mentally simulating an action activates the same neurons in the premotor cortex as performing this action. On the other hand, a few pieces of research suggested "Perceiving is for doing". For example, the same area of monkeys' premotor cortex was activated when they saw an action as when they performed the same action (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Rizzolatti & Arbib, 1998). The Motor Evoked Potentials (MEPs) patterns were the same when they observed an experimenter grasping an object as when they were grasping the object themselves (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). These results suggested perception primes or activates the behavioral tendency itself.

The Common-Coding Hypothesis proposed by Prinz (1990) suggests that language comprehension and language production, or more general perception and action share the same representational systems. That is, the coding system for perceiving behaviors in others and performing those behaviors is the same system. Thus, activation of the mental representation of an action leads to the performance of this behavior, so that people have a natural tendency to imitate (Greenwald, 1970; Wheeler, 1966). This tendency is a consequence of the shared representational systems of perception and action, which does not require motivation or strategy.

Chartrand (2005) further suggested that the perception-behavior link should be bi-directional linked because of the overlap of the representational system of perception and behavior. For example, performing a particular behavior activates the corresponding behavioral representation or schema, which then spreads to the overlapping interpretational schema (Berkowitz, 1984; Carver, Ganellen, Froming, & Chambers, 1983). Perceiving behaviors in others creates the tendency to imitate. And imitating others' behavior, in turn, creates differential interpretation as well.

2.2.3.4 Mirror Neuron and Associative Sequence Learning

Evidence from neuroscience further supported the representation overlapping hypothesis. In the early 1990s, Pellegrino et al. (1992) reported the mirror neuron system (Rizzolatti & Craighero, 2004) in the brain of macaque monkey which fired both when the monkey did a particular action and when it observed another individual making a similar move. This research suggest a functional neural link between perception and action in the monkey's brain (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Later studies suggest a similar system as the mirror neuron system exists in the human brain (Gazzola, Aziz-Zadeh, & Keysers, 2006; Iacoboni et al., 1999; Molenberghs, Cunnington, & Mattingley, 2012; Pobric & de C. Hamilton, 2006; Rizzolatti, Fogassi, & Gallese, 2001; Tranel, Kemmerer, Adolphs, Damasio, & Damasio, 2003), and some action execution neurons act as inhibitor preventing imitation and helping self/other discrimination (Keysers & Gazzola, 2010; Saygin, 2007). However, there are also some concerns about the paradigm of the human mirror neuron system. Different from the studies on monkeys' brains, human brain research typically using EEG (electroencephalogram) or fMRI (functional magnetic resonance imaging) scans, which cannot directly detect the activity of an individual mirror neuron. Instead, researchers have recognized the events from the same neuron region when a person executes an action and when the person perceives another individual doing the same action (Dinstein, Thomas, Behrmann, & Heeger, 2008; Keysers & Gazzola, 2010; Pascolo & Budai, 2013).

Heyes (2001), on the other hand, proposed the theory of Associative Sequence Learning (ASL). ASL suggests that mirror neurons develop by learned associations between observed and execute motor actions in the course of individual development (Cook, Bird, Catmur, Press, & Heyes, 2014; Heyes, 2010). The mirror neurons come from sensorimotor experience, which is mostly obtained through interaction with others. This phenomenon indicates that automatic imitation is a product and process of social interaction and learning (Heyes, 2011). The chameleon effect and behavioral mimicry thus can be regarded as a by-product of the sensorimotor links established through associative sequence learning.

2.2.4 Summary

In this section, the phenomenon of behavioral mimicry is briefed. People would perform the same action that they perceived nonconsciously and unintended. People mimic almost everything that can be observed, which leading to a series of prosocial consequences. The timing used in behavioral mimicry studies varies from 1 to 10 seconds, however, no evidence is given for a certain timing used. The chameleon effect and the perception-behavior link is detailed discussed. Research over decades suggested that behavioral mimicry caused by the shared representations of perception and behavior in the human brain. However, few study has looked into how would contextual variable affect the perception-behavior link. It is suggested that behavioral mimicry is supported by the mirror neuron system and possibly developed through social interaction and learning. Nevertheless, it is unclear that whether this learning for mimicry is limited by genetic factors or not.

2.3 Behavior Research Method

2.3.1 Overview

In the previous section, the studies of behavioral mimicry over the past decades is reviewed. Most of these research utilized confederate as the interface to study behavioral mimicry. However, the recent research started to take advantage of the rapidly developed information technologies, e.g., intelligent virtual agent and immersive virtual environment for mimicry studies (Bailenson & Yee, 2005; Pan & Hamilton, 2015). In this section, these techniques from the perspective of behavior research method is explored. It starts from a brief of ethnomethodology to the use of confederate, then intelligent virtual agent and finally immersive virtual environment.

2.3.2 Ethnomethodology

Ethnomethodology is the study of the knowledge of the common sense of how ordinary people make sense of their everyday world (Hammersley & Atkinson, 2007; Heritage & Atkinson, 1984). It is designed to uncover the norms and behaviors people are using in their daily life. One cannot simply ask a person what norms he or she uses since most people are not wholly conscious of these and are not able to articulate or describe them (Garfinkel, 1964).

When doing ethnomethodological research, researchers would describe the underlying sequential organization of conversation without imposing control on the data collection or a priori hypotheses on the data. For example, some researchers collect data only in everyday settings such as recording conversations on the street (Goodwin, 1985), at the dinner table (Goodwin, 1979). Ethnomethodological research is working primarily with "unstructured data", that is, data is not coded for analysis at the moment of collecting it. Ethnomethodologists take advantage of a small number of cases (even with just one example) in detail. Ethnomethodological research usually involves the analysis of data with an explicit interpretation of the meanings and functions of human actions and produces verbal descriptions and explanations.

Ethnomethodological research could provide abundant, holistic insights into people's views and actions, as well as the nature of the location they inhabit through the collection of detailed observations and interviews (Reeves, Kuper, & Hodges, 2008). Researchers need to record elements, including space, people, activity, object, act, event, time, goal, emotion.

While ethnomethodology is useful to study communicator's behavior in the natural environment, it is not able to test the causal effects of specific phenomena such as behavioral mimicry. Ethnomethodology could be time-consuming. Ethnomethodological data can only be inferred retrospectively. The variation in the collected data is enormous, which makes it difficult to compare different situations or to generalize findings beyond a set of data. Furthermore, collecting dialogue data in an uncontrolled setting cannot be used to make predictions or causal inferences of the mechanisms of the observed behavior (Kuhlen & Brennan, 2013). Some behavior studies need controlled situations. It could be challenging to create such controllability without the use of confederates.

2.3.3 Confederate

The use of confederates is a longstanding tradition in social psychology. This tradition is because, on the one hand, confederates may be the only practical way to collect data for the studies of unusual behavior. Using confederates, on the other hand, may be just a convenience for the experimenter.

In an experiment, a confederate will typically keep their behavior constant and comparable across experimental conditions. In this way, the experimenter can establish a certain degree of experimental control. The same confederate may be used repeatedly through the experiment trails and serve as a sort of stimulus or independent variable. The naïve participant's behavior is then regarded as the target or dependent variable. The confederate's action may be spontaneous or scripted, and it may or may not receive detailed instructions on how to behave (Kuhlen & Brennan, 2013).

Confederates have been used for different purposes, and on top of different communication theories: 1) **confederate was used as the motivational partner** – making the experiment more like a dialogue. According to the social facilitation theory (Zajonc, 1965), the mere presence of an audience improves individual performance. A confederate in this situation is a motivator to turn a participant speaker from the monologue mode to the dialogue mode, or to make a participant listener believe its dialogue partner is another naïve subject. 2) **confederate was used as the egocentric partner** – to establish the two-stage dialogue processes (Keysar, Barr, Balin, & Brauner, 2000) which suggests conversational partners process language in the first stage and adapt to each other in the second stage. In this view, a dialogue is not differ from a monologue. And a confederate's presence as the same effect as a motivation partner. In this case, the listener mainly provides feedback for understanding, and the speaker mainly does corrections or repairs if there is a misunderstanding. 3) **confederate was used as the interactively aligned partner** – to automatically align the mental representations of the conversational partner (Pickering & Garrod, 2004a). It suggests dialogue is fundamentally different from a monologue. The presence of confederate changes the core processes in language production and comprehension. However, it considers the listener need only listen passively and is less important than the speaker. 4) **confederate was used as the collaborative**

partner – to make a coordinated dialogue. This considers conversation as a joint action or collaborative process (Bavelas, Coates, & Johnson, 2002; Clark, 1996; Clark & Wilkes-Gibbs, 1986; Holler & Wilkin, 2011a; Richardson et al., 2007; Richardson & Dale, 2005; Roberts, 1996). In this situation, a listener can actively shape the interaction and act as a co-narrator through verbal or nonverbal feedback (Bavelas et al., 2000). A confederate is required to perform behaviors in a highly contingent and precisely timed way to show mutual understanding and uptake in the grounding process or otherwise break the dialogue.

Using confederates has particular advantages, it can 1) collect data efficiently 2) increase the frequency of rare events 3) reduce exuberant data 4) focus on the individual as the unit of study 5) go beyond monologue 6) adhere to the standard statistical tests 7) reduce complexity (Kuhlen & Brennan, 2013).

On the other hand, the use of confederate may distort the interaction and the processes, representations, and behavior in study because confederates' behavior may be systematically different from naïve subjects' behavior. Bavelas and Healing (2013) reviewed fourteen experiments about the overall rate of gesturing in dialogues in a variety of participants' mutual visibility and found that quasi-dialogues experiments using confederates would produce significantly different results comparing to free dialogues experiments. Kuhlen and Brennan (2013) suggested that the faulty use of confederates may lead to problematic results. Confederate as a listener could make the experience unnatural and invalid if it is prevented from responding or is unable to depart from a script. On the other hand, confederate, as a speaker, could lead to different behavior patterns than an authentic speaker if it ignores addressees' needs for clarification or behaves in inauthentic or unexpected ways. (Kuhlen & Brennan, 2010; Kuhlen, 2010; Schober, Conrad, & Fricker, 2004). Kuhlen and Brennan (2013) further suggested 1) a confederate should be blind to the study design to avoid bias 2) a confederate should not be identified as confederate by subjects 3) a confederate should not know too much about the experiment topic as an listener unless it is required 4) a confederate's behaviors should not be scripted.

Confederates are better used with unusual behavior or low-frequency linguistic forms. However, they are risky to be used as a listener, especially if their nonverbal behavior is uncontrolled, and they know too much about the experiment. In addition to the four concerns noted above, if using a confederate, it is the best to report the details about how confederates are integrated into the experimental protocol and carefully consider whether confederates are needed or not in the experiment (Kuhlen & Brennan, 2013).

2.3.4 Intelligent Virtual Agent

Intelligent virtual agents (IVAs) are intelligent virtual characters that can communicate with humans and other agents using natural human communication modalities such as speech, gestures, facial expressions and movement. They have been used to study human communication and social behavior, along with the development of multimedia technology. IVAs are capable of real-time perception, cognition, emotion, and actions to participate in dynamic social environments. The construction and study of IVAs locate in the field of the interdiscipline of computer science, psychology, cognitive sciences, communication, linguistics, interactive media, human-computer interaction, and artificial intelligence (Beskow et al., 2017; "IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents", 2019).

The research of IVA derives from the Turing Test (originally the "imitation game") (TURING, 1950). The generally known Turing test (standard Turing test) can be described as follow:

> If an interrogator cannot distinguish a human responder from a computer responder by asking them questions, then the computer can be regarded as being intelligent (Traiger, 2003).

Artificial Intelligent reseachers created a kind of computer program called chatterbot (Deryugina, 2010) seems to pass the Turing test. ELIZA developed by Joseph Weizenbaum's in the 1960s (Weizenbaum, 1966), for example, is able to fool some people to believe that they are talking to a real person and it is claimed (though very controversial) to be the first program that passed the Turing test. ELIZA implemented simple keywords pattern matching algorithm (stimuli-response pattern matching) to respond to the user's input. It can simulate intelligent behavior without using much intelligent or knowledge about the world or language. Chatterbots are continuely evolving with the developing of the natural language processing techniques. They may be implemented with rule based dialogue system, statistical model (machine learning) based dialogue system or state-of-the-art deep learning based dialogue system, and become inceasingly intelligent.

The verbal version of the Turing test also generalizes naturally to all of human performance capacity, verbal as well as nonverbal (Oppy & Dowe, 2003). Embodied conversational agents (ECAs), for example, are virtual characters driven by artificial intelligence to perform verbal and nonverbal communication. They have multimodal interface including speech, facial expression, gestures, postures, etc. They are integrated with dialogue systems for verbal and nonverbal communication (Cassell, Sullivan, Churchill, & Prevost, 2000). They are designed to be flexible and versatile in communicating with real humans and facilitate a high sense of presence, co-presence, and social presence to elicit natural behavior in real humans (De Leo, Diggs, Radici, & Mastaglio, 2014).

Researchers tried different methods to improve the sense of social presence or co-presence of an IVA through modifications to its behavior during an interaction (Garau, Slater, Pertaub, & Razzaque, 2005; Huang, Morency, & Gratch, 2011; Oh, Bailenson, & Welch, 2018). For example, participants reported a higher sense of co-presence of the IVA with head movement compared to the IVA without head movement (Bailenson, Beall, & Blascovich, 2002).

IVAs have been explored in many places. For example, researchers used IVAs to study **eye gaze in communication** (Bee, André, & Tober, 2009), **interpersonal distance** (Bailenson, Blascovich, Beall, & Loomis, 2003), **emotion for decision making** (de Melo, Gratch, & Carnevale, 2015), **Proteus Effect** (Peña, Hancock, & Merola, 2009; Yee & Bailenson, 2007), **language alignment** (Bergmann, Branigan, & Kopp, 2015).

IVAs are used a lot in language studies. For example, Bergmann et al. (2015) suggested language behavior changes with/without the presence of an IVA. The social cues and presence created by IVA activated automatic social reactions. Participant adapted their behavior based on their expectation of the system (Pearson, Hu, Branigan, Pickering, & Nass‘, 2006). Furthermore, Heyselaar et al. (2017) suggested using a human-like IVA is identical to using a real human in language behavior in a dialogue.

In the experiments on human interaction, IVAs can provide a significant advance toward the use of confederates producing scripted behaviors (Bavelas & Healing, 2013; Kuhlen & Brennan, 2013). In real humans, different nonverbal behaviors are often highly correlated with each other. Altering real human's behavior cannot avoid bringing in other uncontrolled factors from behavior shift. In principle, agent behaviors are entirely controllable and always blind to experimental manipulation. Controlling some of the actions in virtual humans allows us to maintain complete independence among these behaviors (Bailenson, Beall, Blascovich, Raimundo, & Weisbuch, 2001).

Nevertheless, people do not treat an IVA precisely as they would treat a real human. For example, participants showed less engaged, sincere, interests, and had a more poor attitude towards an IVA than an actual human pretending to have the same symptoms (Raij et al., 2007). IVA's advice was more rarely sought out compared to a physically present robot (Pan & Steed, 2016).

The intelligent virtual agent used to be presented with monitors. Recently, an increasing number of research combined the use of intelligent virtual agent with virtual reality (Bailenson et al., 2003; Daher et al., 2017; Heyselaar et al., 2017; Pejsa, Gleicher, & Mutlu, 2017). The rich information and isolation which virtual reality can provide make it a game-changer for behavior research. This is discussed in the following section.

2.3.5 Immersive Virtual Environment

Virtual Reality systems can be presented in different ways, for example, via head-mounted displays (HMDs), Cave Automatic Virtual Environment (CAVE) (Cruz-Neira, Sandin, DeFanti, Kenyon, & Hart, 1992), projectors, or desktop screens. Immersive Virtual Environment (IVE) (aka Immersive Virtual Reality, IVR) is typically experienced in an HMD. The HMD provides an immersive experience (Slater, 2009) with (1) 3D stereo vision (2) isolation from the physical world and (3) responsive motion tracking. An IVE is one that perceptually surrounds the users, increasing their sense of presence or being within it. It has two features: first, the user's movements are tracked and automatically and continually updated in the virtual environment; second, it enables the construction of a variety of scenarios and tasks.

IVE can provide a high ecological validity (Schmuckler, 2001), reproducibility and high experimental control. It was widely used for the studies in, e.g., **therapy** (Riva et al., 2010), **social neurosciences** (Bohil, Alicea, & Biocca, 2011; Parsons, 2015), **psychology** (Loomis, Blascovich, & Beall, 1999), **language studies** (Peeters, 2019), **social interaction** (Pan & Hamilton, 2018). Moreover, IVE enables researchers to experiment with impossible scenarios such as asking questions that might otherwise be limited by ethical concerns (Slater et al., 2006), doing physically dangerous tasks (Bhagat, Liou, & Chang, 2016).

IVEs provide researchers the ability to filter a multitude of cues that contribute to any given experience, which is difficult in the physical world (Bailenson, Blascovich, et al., 2001; Hale & Hamilton, 2016; Healey, Frauenberger, Gillies, & Battersby, 2009). This ability makes it possible to carry out controlled manipulations of the user's experience (Bailenson & Yee, 2003). They also give researchers access, in principle, to all of the participant's motion data, which can provide useful additional dependent variables for analysis (Blascovich et al., 2002).

2.3.5.1 Motion Tracking

Motion tracking is one of the critical features of an IVE. Head motion tracking is the fundamental function of the IVE to update the visual display and present information on the direction of a participant's attention. Most low-end HMDs, e.g., Cardboard VR or Oculus Go, provide primary motion tracking function in 3 degrees of freedom (DoF) for head movements with a gyroscope sensor. These HMDs can only present virtual environments from a fixed viewpoint. The wearers can rotate their heads while adjusting the head position is not enabled. The higher-end HMDs can provide 6 DoF motion with head orientation and position tracking. These HMDs usually come with extra tracking module, for example, motion cameras for Oculus Rift, lighthouses for HTC VIVE. Those HMDs could further provide hand tracking with controllers yet with a limited range of movement (ROM). More advanced HMDs, for example, Oculus Rift S and Oculus Quest, can do an inside-out tracking so that their ROM is unlimited. There are also extensive motion tracking systems for HMDs, such as eye-tracking for gaze interaction, finger tracking (e.g., Leap Motion, data gloves) for gesturing, and full-body motion tracking (e.g., Vicon optical motion capture system) enables the best natural communication in the IVEs. It is believed that the rich capture of human behavior is essential for social interaction research.

Motion tracking in the IVE has been used in many ways, including: 1) generating realistic yet well-controlled virtual character animations stimuli (De La Rosa, Ferstl, & Bülthoff, 2016) 2) generating real time realistic responsive interactions between the participant and other objects or characters – this is widely used in gaze interaction studies (Beall, Bailenson, Loomis, Blascovich, & Rex, 2003; Fallis, 2013; Fox & Bailenson, 2009; Garau et al., 2003), social distance studies (Bailenson et al., 2003; Fallis, 2013; Pan, Gillies, Barker, Clark, & Slater, 2012), mimicry studies (Bailenson & Yee, 2005; Bailenson et al., 2007; Hale & Hamilton, 2016; Verberne et al., 2013) 3) record natural and unconstrained behaviors for psychological measurements such as proxemics (Bailenson et al., 2003; McCall & Singer, 2015), approach as a measure of trust (Hale, Payne, Taylor, Paoletti, & Hamilton, 2018), and imitation (Pan & Hamilton, 2015).

2.3.5.2 Avatar

A critical part of the IVE experimental paradigm is the use of avatars to represent people within the IVE. An avatar is a virtual character directly controlled by a real human. It is possible not to use avatars in VR. However, avatars enable an embodied experience for participants and enhance the body-ownership illusion (Bergström, Kilteni, & Slater, 2016; Maselli & Slater, 2013). This is achieved with the multi-sensory integration of IVE (Ehrsson, 2007; Lenggenhager, Tadi, Metzinger, & Blanke, 2007; Slater, Spanlang, Sanchez-Vives, & Blanke, 2010). The sensory illusions created with avatars in the IVEs produce a more or less believable simulation of reality and a sense of 'presence' in the IVE (Biocca & Levy, 2013; Slater, 2009). They are important for the effectiveness of VR training in medical, military, educational simulations, as well as for therapeutic applications (Slater & Steed, 2000; Slater, Usoh, & Steed, 1995).

There are two different use cases for avatars – self-avatars and other-avatars. Self-avatars are the avatars being embodied by the participants. They are used mostly in the studies of personal psychological effects. Compared to traditional perspective-taking, which relies on imagination, this embodied experience of self-avatar puts participant directly into another person's shoes, led to greater self-other merging, favorable attitudes, and helping towards persons with disabilities. For example, people with normal color vision experienced being color blind, and in turn, changed their subsequent behavior toward others (Ahn, Le, & Bailenson, 2013). People became more environmental conservative after experienced cutting a virtual tree in IVE (Ahn, Bailenson, & Park, 2014). People became less racial biased after being embodied into avatars with black skin (Yee & Bailenson, 2007). Self-avatars also help spatial cognition and navigation in IVEs (Bohbot, Lerch, Thorndycraft, Iaria, & Zijdenbos, 2007; Driscoll, Hamilton, Yeo, Brooks, & Sutherland, 2005). Compared to explore the IVE without an avatar, people embodied with a full avatar make less error in distance judgment (Mohler, Bülthoff, Thompson, & Creem-Regehr, 2008). Other-avatars, on the contrary, are the avatars being embodied by the other people that the participant is interacting with. Other-avatars are usually used in the studies of social interaction. Using avatars can provide better controllability than human confederate and much more natural social cues such as facial expression, gestures and postures.

A key question for IVE based experiments is whether the avatar/agent and virtual environment are realistic enough to elicit the participant's typical responses. The most obvious level at which this issue arises is visual appearance. As it is suggested, virtual bodies are sufficiently realistic to induce the body ownership illusion. Additionally, seeing a realistic virtual body in the same location and posture as the physical body also engenders a sense of body ownership (Maselli & Slater, 2013). Latoschik et al. (2017) studied the impact of avatar realism on the experience of embodiment and quality of social interaction. They compared a neutral abstract avatar representation (wooden mannequin) with high fidelity scans of real humans. The results suggest that the realistic avatars were significantly more human-like when used as avatars for the others and more accepted in terms of virtual body ownership.

The issue is not necessarily one of visual realism. Argelaguet (2016) studied hand ownership in high and low appearance realism. In this study, participants were embodied in an avatar with hands tracked with Leap Motion. They were asked to put their virtual hand close to a virtual spinning saw. With realistic hands, participants tended to spend more time to perform the task and became more careful than with unrealistic hands. This research suggested that a realistic avatar would create a higher sense of ownership. However, the survey results showed that the feeling of agency (sense of controllability of one's actions) was more robust with low realistic hands. This may be due to a trade-off between visual realism and tracking accuracy. The lack of agency might be related to the effect of the uncanny valley (Mori, MacDorman, & Kageki, 2012), which suggests humanoid objects imperfectly resemble actual human beings provoke uncanny or strangely familiar feelings of eeriness and revulsion in observers. This study implies that although a realistic avatar created a higher sense of ownership, it requires fine-grained motion mapping for the feeling of agency.

More recently, Zibrek et al. (2019) compared the realistic, place illusion, and social presence between the photorealism virtual character to stylism rendered virtual character and a simple virtual character. In the study, participants just needed to listen to the virtual character and do a survey. Their result suggested that although the rating of visual realism does not differ from conditions, the evaluation of movement realism is higher with the stylism rendered virtual character than the photorealism virtual character. This result is consistent with Argelaguet's finding, which suggested the trade-off between visual realism and tracking accuracy. Zibrek's study further indicated a stronger place illusion with the photorealism virtual character. At the same time, there was no difference in social presence with a virtual character in a different render style.

Realism may also be linked to persuasiveness. Guadagno (2007) compared the influence of virtual agents that performs no nonverbal behavior with an agent that makes complex nonverbal behaviors such as tracking participants' eye contact. Their result suggested that realistic behavior is more influential than non-realistic behavior. A later study also indicated the importance of realistic nonverbal behavior. With more nonverbal cues, the prime effect was more significant (Heyselaar et al., 2017). It pointed out the importance of motion realism, possibly over visual realism, for maintaining the sense of immersed engagement in IVE experiments.

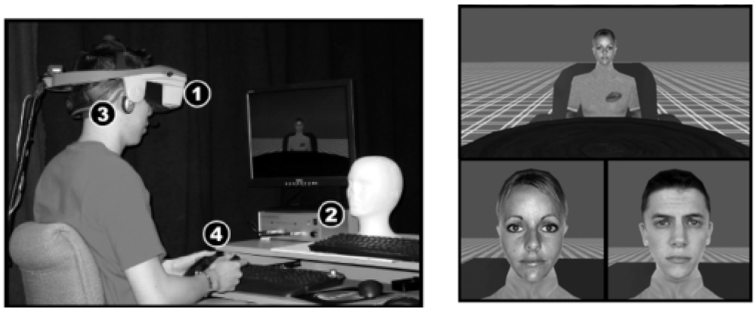
2.3.6 Summary

In this section, a list of different methods in behavioral research is compared. Ethnomethodology can provide rich and holistic insights into natural behavior. It is also time-consuming and challenging to generalize findings. Confederate is fit for laboratory behavior research. It is more practical to collect data. However, it cannot avoid bringing bias from itself. The intelligent virtual agent is the recently advanced technique. It is perfect for replacing confederate since it is entirely controllable and bias-free. Although it is believed that IVAs can eventually behave as the same as real humans. However, there are concerns raised with the imperfect/unrealistic behaviors of the IVA. These imperfect/unrealistic behaviors will lead to unnatural or different attitudes when interacting with IVAs. Finally, an avatar with proper motion tracking in the IVE can be an excellent alternative to IVA for its improved, realistic movement. It is a promising method to be used in future behavior research.

2.4 Digital Chameleons

The term "Digital Chameleons" was coined by Bailenson and Yee (2005) and built on the ideas of the chameleon effect. In the study of the chameleon effect, Chartrand and Bargh (1999) claimed that people automatically mimic each others' movements and behaviors unconsciously during an interaction, usually within a short window of time between three to five seconds. This result leads to the prediction that if one person simply repeats aspects of another person's movements, this ought to enhance their credibility and persuasiveness for the other participant. The degree of mimicry must, of course, be carefully controlled to avoid creating a sense of parody or irony. Nodding is ostensibly the right candidate since it is a relatively unmarked and positively valenced behavior.

Bailenson and Yee (Bailenson & Yee, 2005) tested this prediction using an intelligent virtual agent who delivers a message that aims to convince students that they should always carry their ID cards. Participants were seated and wore a Head-Mounted Display and saw an agent in the IVE (Figure 2.2). The message was delivered to a seated participant through a Head-Mounted Display. After the speech, their post-hoc ratings of the persuasiveness of the speech by the agent were measured by questionnaires. The experiment compared the effectiveness of the agent in two different conditions: mimic and non-mimic. In the mimic condition, the agent reproduced exactly the head movements of each participant with a fixed 4s delay. While in the non-mimic condition (corresponding to the Playback condition in this thesis), the agent's head



(a) (b)

Figure 2.2: The experimental set-up for the settings of the 'Digital Chameleons'. Adapted from Bailenson, J. N., & Yee, N. (2005). Digital chameleons. *Society*, *16*(10), 814–819

movements were controlled by a canned sequence recorded from the head movements of the previous participant. The results suggested that the mimicking agents were more persuasive than the non-mimicking agents. Furthermore, motion analysis suggested that the overall movements of men are significantly larger than women.

In a second study, Bailenson and Yee (2007) found a relationship between liking and mimicry using a mechanical device that mimicking participants' handshakes. They suggested that males respond more strongly to mimicry than females. Verberne et al. (2013) used virtual agents to mimic participants' head movements during a route planner game and investment game. The results showed that the mimicking agents were more liked and trusted by participants in the route planner game rather than the investment game. They suggested that the "Digital Chameleons" effect comes with conditions: first, the effect might need a certain time to be obtained; second, the type of behavior might be a factor, i.e., might depend on the consequence of the behavior being predictable. Hasler el al. reported that participants that were postural mimicked by socially/ethnically conflicted IVA showed greater empathy than participants that were not mimicked (2014). Stevens et al. (2016) studied the likability of an agent that mimicked a participant's head nods, or eyebrow raises when the agent repeated a participant's sentence. They suggested that the likability of the agent depended on mimicry and the prominence of visual cues moderated it. The more prominent signals the participant produced, the higher the judged lifelikeness of the agent in the mimic condition.

However, the "Digital Chameleons" effect has not been consistently replicated. Riek et al. (2010) compared human-robot rapport in three conditions: the robot mimics participants' full head movement, the robot mimics participants head-nodding only, and the robot does not mimic participants. They found there is no significant difference in the human-robot rapport between the three conditions. They suggested that the survey measure might be unsuitable for rating human-robot interaction. They also noted that men has more significant movements than women and that participants' actions were possibly unexpected. In an especially careful study, Hale et al. (2016) extended mimicry to head and torso movements, and measured the rapport and trust towards the agent after participant and agent had carried out a photo description tasks. The results suggested that mimicry has no significant relationship with human-agent rapport or trust. Hale et al. further suggested the positive social effects of being mimicked may be more subtle or fragile than is generally assumed. Some critical behaviors which create the natural mimicry might be absent with the intelligent virtual agent.

2.5 Summary

Behavior research is complex and it is facing the replication crisis (Bardi & Zentner, 2017; Diener & Biswas-Diener, 2016; Earp & Trafimow, 2015; Hantula, 2019; Lilienfeld, 2017; Maxwell, Lau, & Howard, 2015; Tincani & Travers, 2019). Researchers frequently do not find the same positive result when replicating past study. One reason for this might be the original result is false postive. Since researchers have lots of degrees of freedom (as many as at least of 34 DoFs) in planning, running, analyzing and reporting of psychological studies, there would be many ways to obtain statistical significant effect, for example, reporting subsets of experimental conditions (Simmons, Nelson, & Simonsohn, 2011; Wicherts et al., 2016). It is important to do replication studies even for many times to sort out the true positives. The "Digital Chameleons" study has not been consistently replicated over the decade. It is an interdiscipline study mixed research fields such as speaker and listener coordination, head nodding, behavioral mimicry, confederate, intelligent virtual agent, and immersive virtual reality. The study could be easily messed up if it is not carefully designed. Or it could create an unnatural situation which would not happen in the real world. Thus, we had a holistic review of the important research related to these fields in this chapter. A few concerns about experimenting with nonverbal behavior were discussed. In chapter 7, how does the "Digital Chameleons" study encounters these problems is further discussed.

Chapter 3

Methodology

This chapter describes the methodological approach for the research. It starts from the rationales of the replications of the "Digital Chameleons" study, followed with a description of the evaluation methods of the effectiveness of the virtual agent/avatar. In the experiments, fine-grained motion data is collected for the study of head nods coordination. the techniques for the analysis of motion time-series data is explained in the last section.

3.1 Replications of “Digital Chameleons”

In chapter 2, the literature of behavioral mimicry and “Digital Chameleons” studies is reviewed. Behavioral mimicry is believed to be nonconscious and automatic behavior. People automatically mimic others would be more socially influential (Chartrand & Bargh, 1999; Chartrand & Lakin, 2013). Bailenson and Yee designed the “Digital Chameleons”. They claimed that the IVAs automatically mimic participant’s (listener) head movements are more persuasive than the IVAs nods the same as their previous participant.

However, the research on head nods and speaker-listener coordination, noted in Section 2.1.1, suggested that the speaker and listener nod differently, and their behaviors are naturally coordinated. It is possibly invalid to compare the effectiveness between the IVAs mimicking a listener’s head nods and the IVAs using another listener’s head nods. Their head nods are likely to depart significantly from a speaker’s natural head nods. These head nods are timed in a way that is out of step with normal dynamics of a conversation exchange since the nods occur independently of the content. The following failed replications of the “Digital Chameleons” experiment further throw doubt on this paradigm.

The discussion on the reviewed literature suggests that the “Digital Chameleons” paradigm ought to fail if not considering the coordination of head nods in dialogue. That is, the effectiveness of the IVAs that is mimicking a listener’s head movements, should have no difference from the IVAs that is using another listener’s head movements. Furthermore, their effectiveness should be no difference, even comparing to the IVAs that uses a recorded animation of a real speaker making the speech of the same content. On the one hand, the IVA using recorded animation or algorithms to drive behavior could quickly lose track of the coordination between speaker and listener. On the other hand, using an avatar speaker instead of an IVA speaker may lead to different results. This is because avatars can deliver natural interaction with a full-body motion capture system.

Therefore, two experiments are designed to partially replicate the “Digital Chameleons” study. An IVE is built for participants interacting with an IVA or another participant embodied in an avatar. The architecture of the system is described in Chapter 4.

The first experiment is designed to replicate the original “Digital

Chameleons" study and further test the effectiveness of an IVA using recorded speaker's movements. It recreates the two original conditions: 1) Mimic – IVA copies the participant's head movements with a 4s delay; 2) Playback – IVA uses head movements from playback of a previous participant and therefore disconnected with the content. A third condition is added: 3) Recording: IVA uses the prerecorded speaker's movements. The IVAs deliver a persuasive message to the participants with the head movement manipulation based on the condition it is in, and other variables remain the same. In this way, the original "Digital Chameleons" effect is tested. Besides, with the third condition, it further examines the effectiveness of the IVA using a real speaker's movements that is decoupled from the real listener's movements.

The second experiment is designed to test the difference of the effectiveness between the mimicking speaker and the natural speaker with the "Digital Chameleons" paradigm. Virtual avatars are used to replace IVA speakers. The virtual avatars are driven by the real-time motion-captured data of one participant who is asked to deliver the persuasive message to the other participant. In this experiment, four conditions are created: 1) Mimic – the virtual speaker copies the listener's head nods at 4 seconds delay; 2) Playback – the virtual speaker uses the previous listener's head nods; 3) Natural – the virtual speaker uses exact the real speaker's head nods; 4) Recording – the virtual speaker uses a prerecorded animation of head nods.

The experiments test the effectiveness of the virtual speakers in the same way as the "Digital Chameleons" (Bailenson & Yee, 2005) (Chapter 5). At the same time, head movement data is collected to analyze the coordination of head nods in these experiments (Chapter 6).

3.2 Evaluating the Effectiveness of Virtual Speaker

Two methods are used to evaluate the effectiveness of the virtual speaker. To assess the "Digital Chameleons" paradigm, the same evaluation method as the original study is used. Besides, principal component analysis (PCA) is used to reassess it.

3.2.1 Measures for Effectiveness

The same questionnaires in the "Digital Chameleons" study is used to measure the effectiveness of the virtual speaker. The questionnaires have 4 agreement questions, which asked participants to assess the ID card regulations: "I agree with the plan to implement ID cards", "I think the proposed ID cards are: valuable, workable, needed"; 12 questions which asked their impression of the virtual presenter: "The presenter was: friendly, likable, honest, competent, warm, informed, credible, modest, approachable, interesting, trustworthy, sincere"; and 7 questions which asked them to assess the social presence of the virtual presenter: "To what extent you: enjoyed the experience, want to meet him/her again in current situation, feel him/her isolated, want to meet him/her again, feel comfortable with him/her, feel him/her cooperative, feel self-conscious or embarrassed with him/her". The answers for the questionnaires are Likert scale range from 1 strongly disagree to 7 strongly agree. There are also four separate open-ended paragraphs concerning their experience in virtual reality: "Please list any thoughts you may have about the interaction with the presenter", "Was there anything unusual about this interaction?"; and the agent's movements: "Please write a few sentences about the presenter's LIP movements while speaking", "Please write a few sentences about the presenter's HEAD movements while speaking" (see Appendix C).

The composite measures replicates the measurements of the IVA's effectiveness in the "Digital Chameleons" study by taking the mean response to the 4 agreement questions as the measurement 'Agreement', the mean response to the 12 impression questions as the measurement 'Impression', the mean response to 7 social presence questions as the measurement 'Social Presence', the mean response to the all 23 items as the measurement 'Effectiveness'.

According to the central limit theorem, the data composed from multiple factors should form a normal distribution. Thus, the measurements of 'Agreement', 'Impression', 'Social Presence', 'Effectiveness' should form 4 normal distributions. The difference in these measures is tested between groups in different conditions with linear statistical methods, such as ANOVA, GLM, GLMM, and Student T-test for pair-wise tests.

3.2.2 Principle Component Analysis

The composite measures of the agent's effectiveness used in the "Digital Chameleons" study are arbitrarily created. A better way to compose the measurement is to do a factor analysis or PCA to generate the compositions. The validity of the generated compositions is tested with Cronbach's alpha. Only the composition with an alpha value greater than 0.7 is considered to be valid.

Same as noted above, the difference in these compositions between conditions is tested with linear statistical methods.

3.3 Head Nods Analysis

The purposes of replicating the "Digital Chameleons" study are 1) testing the initially reported effect; 2) collecting head nods motion data to test the coordination of head nods in dialogue. This section describes the measurements for head nods and their analysis methods.

3.3.1 Descriptive Statistics of Head Nods

The first step to analyze head nods is to get an overview of head nods in the dialogue. This can be achieved by looking at the descriptive statistic data of head nods. All the head nods time-series data is collected with the Vicon optical motion capture system. The standard deviation, maximum, and total degrees of head nods are calculated for the data. These measures could help us to understand the form of ordinary head nods, for example, the distribution of ordinary head nods.

The straight-ahead position (i.e., looking directly to the front) is always calibrated as the origin for head movements. Since the raw data is the head rotation recorded in degrees range from 0 to 360 and the rotation of head at the straight-ahead position is 0 degree, the time-series will be discontinued at the straight-ahead position, i.e., jumping between 0 and 360 degree. To solve this problem, the data is adjusted, so that the data is mapped to the range of -180 to 180 degrees and the mean degree of head nods in the time series should be always near 0.

The standard deviation of the degrees of head nods varies with each participant and tells the range of degrees that most of the head nods lie in.

The maximum degree of head nods is the maximum value (in degrees) of deviation from the straight-ahead position. It suggests the upper movement range of head nods in degree. It would serve as an approximation of visual attention. If the participant lost interest or get bored with the virtual speaker, he will look away from the virtual speaker, for example, look around (look at the floor or ceiling) in the IVE during the dialogue. Thus, the maximum degree of head nods will increase.

The total head nods is the overall journey the head traveled in the head pitching direction. The total degrees of head nods would show the participants' listening status. The participant would move less as a listener than a speaker. The total degrees of head nods can be calculated with the equation:

$$totalNods = \sum_{t=0}^{T} |x_t - \bar{x}| \tag{3.1}$$

where T is the length of the head nods time-series. x_t is the degrees of the head nods at time t. $\bar{x}$ is the mean degrees of the head nods over the time series.

This approach is quite similiar to the motion energy detection (MED) method (Davis & Bobick, 1997; Grammer, Honda, Juette, & Schmitt, 1999). The motion energy detection method is widely used for the assessment of quantitative description of nonverbal behavior. It is applied to the recorded video of the nonverbal behavior. The MED produces the total amount of change that occured during a certian time span by substracting successive video frames from each other, while this method accumulates the total amount of head movement change in the pitch direction. However, since the data is collected from the fine-grained motion capture system, this method is much more precise than the MED and works irrespective of the orientation of the participants.

3.3.2 Head Nods by Frequency

The descriptive statistic data gives an overview of head nods in the dialogue. Previous research suggests people communicate with high-and-low-frequency head nods (Hadar et al., 1985). Listeners produce more high-frequency head nods than speakers (Hale et al., 2019). It is necessary to test head nods by frequency for the groups in different conditions as well. In the previous research, head nods are tested from 0 to 6.5 Hz

(Hale et al., 2019). In this study, 0 to 8 Hz is chosen to cover the high and low frequency head nods.

Frequency analysis can usually be done with Spectral Analysis, Fourier Transform Analysis, or Wavelet Analysis. However, these methods are too complicated in this situation where this studycomparing the number of head nods within the different frequency ranges. Amore straightforward method to analyze head nods by frequency is proposed:

Peaks in the head nod time-series are treated as the point that the participant change the direction of head movement and counted as a nod. Those peaks can be easily detected with the MATLAB function 'findpeaks()'. 'findpeaks()' has many parameters such as minimal peak height, minimal height difference, minimum peak width, etc. The default parameters are used in this study. With the default parameters, the 'findpeaks()' function would regard all the spikes in the time-series as peaks since this study is going to check the effect of subtle head nods. The total amount of head nods is counted as the number of peaks in the head nods time-series data. For our customized frequency analysis, a low pass filter is used on the time series data with the cutoff frequency set to increase slowly from 0 to 8 Hz in the resolution of 0.1 Hz. This method is similiar to Hale's nod detector (Hale et al., 2019) which counting the frequncy of the signal makes a zero crossing in a second as the head nodding frequncy. However, it does not require the signal to make zero crossing to be counted as a head nod. Thus, subtle head movements are detected. It would also be expected that more high-frequncy head nods are detected with this method than Hale's. With the recorded head nods time-series data, the difference of the number of head nods between the speaker and listener is tested with the paired t-test in the frequency range 0-8 Hz.

3.3.3 Head Nods Coordination

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.

This study uses Cross Recurrence Quantification Analysis (xRQA) method to test the coordination of speaker and listener head nods in dialogue. It is an extended method of Recurrence Quantification Analysis (RQA).

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

xRQA reconstructs two one-dimensional time-series data to pairs of points in a higher Embedding Dimension phase space using Time-Lagged copies (Takens, 1981). It calculates the distances between the reconstructed pairs of points. The points pairs that fall within a specified distance (Radius) are considered to be recurrent. The recurrent points are visualized with Recurrent Plots (RPs). The RP shows the overall amount of repetition of (%REC), the longest sequence of repeated behaviors (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.

RQA takes seven parameters as inputs including: (1) Embedding Dimension (m), (2) Delay (τ), (3) Radius, (4) Rescale, (5) Norm, (6) Range, (7) Line. The outputs of RQA are the recurrent plot (RP) and quantitative results %REC, %DET, and LMAX. The parameters should be set based on certain principles as below:

The *Embedding Dimension* (m) needs to be chosen to maximize the information of the system, usually estimated by the nearest-neighbor methodology or False Nearest Neighbor (FNN) analysis (Kennel, Brown, & Abarbanel, 1992). This methodology works well on stable and low-noise systems, such as the Lorenz attractor (Lorenz, 1963). However, with real-world data, noise, and non-stationarities (transients, drifts) in the system modulate the critical embedding dimension. Thus, in practice, embedding dimensions of 10 to 20 is typically used with biological systems, but no higher. If the embedding dimension is set to too high, even random/stochastic systems (which in theory exhibit recurrence only by chance) display strong, yet artifactual, patterns of recurrence.

The *Delay* (τ) should be selected to minimize the interaction between points of the measured time series. Two common ways of selecting a proper delay include finding the first zero in the autocorrelation function (Priestley, 1988) or the first minimum in Average Mutual Information (AMI) function (Prichard & Theiler, 1994) of the continuous time-series. There are two criteria for the selection of delay 1) the delay should be large enough so that the various coordinates contain as much new information as possible; 2) the delay should be small enough that the various coordinates are not entirely independent. However, according to Grassberger, the two criterions mentioned above could only be applied to the time-series of which the embedding dimension is 2. There is no optimal choice of delay, and short delays will, e.g., enhance the effect of noise on dimension estimates and improve the efficiency of noise reduction algorithms. Thus, if one is willing to apply noise reduction before estimating the embedding dimension, short delays will be better (Grassberger, Schreiber, & Schaffrath, 1991). In special cases, the delay can also be set to 1 for continuous flows, if the goal is to perform waveform matching (recurrence matching of similar waveforms, point for-sequential-point). For example, the RQA parameters are set to $m = 10, \tau = 1, radius = 50$ in a study of cardiac autonomic nervous system (Takakura et al., 2017).

The *Radius* is always expressed in units relative to the elements in the distance matrix. In effect, the radius parameter implements a cutoff limit (Heaviside function) that transforms the distance matrix (DM) into the recurrence matrix (RM). All (i, j) elements in DM with distances at or below the RADIUS cutoff are included in the recurrence matrix (value = 1), but all other elements are excluded from RM (value = 0). There are three guidelines for selecting the proper radius 1) Radius must fall with the linear scaling region of the double logarithmic plot; 2) %REC must be kept low (e.g., 0.1 to 2.0%); and 3) Radius may or may not coincide with the first minimum hitch in %DET.

The *Rescale* option controls the rescale of the DM by dividing down each element in the DM by either the mean distance or maximum distance of the entire matrix.

The *Range* is defined by the selected starting point and ending point in the time series to be analyzed. In effect, the range defines a window on the dynamic under investigation.

The *Norm* has three options: minimum norm, maximum norm, and Euclidean norm. As implied by its name, the norm function geometrically defines the size (and shape) of the neighborhood surrounding each reference point.

The last parameter is termed the *line* parameter. This parameter is essential when extracting quantitative features from recurrent plots , but exerts no effect on the recurrence matrix itself. If the length of a recurrence feature is shorter than the line parameter, that feature is rejected during the quantitative analyses. Typically, the line parameter is set equal to 2 because it takes a minimum of two points to define any line. But it is possible to increase the line parameter (in integer steps) and thereby implement a quantitative filter function on feature extractions, but this is not necessarily recommended.

When comparing across people or groups, fixed parameters are preferred (Gray, 2016).

The first thing to be done to perform RQA, is phase space reconstruction (Takens, 1981). Phase space reconstruction is based on the fact that one dimension time-series is embedded with higher dimensional information. Thus, the entire multi-dimensional dynamics of a system can be extracted from a one-dimensional time-series data.

The algorithm is demonstrated with Lorenz Attractor (Lorenz, 1963) as well as the real head nodding data:.

The Lorenz Attractor can be generated with the Lorenz equations:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y, \\ \frac{dz}{dt} &= xy - \beta z. \end{aligned} \tag{3.2}$$

in which $\sigma = 10, \beta = 8/3, \rho = 28$. Figure 3.1 illustrates the XYZ time series of the Lorenz Attractor and its trajectory in the 3 Dimension Space[1].

The complete information of the Lorenz Attractor is presented at the moment (Figure 3.1). And the complete head movement is illustrated with the figure 3.2. However, in the real world, the complete information

[1]The program for this visualization has been made open source at https://github.com/Leshao-Zhang/Visualisation-Cross-Recurrence-Quantification-Analysis

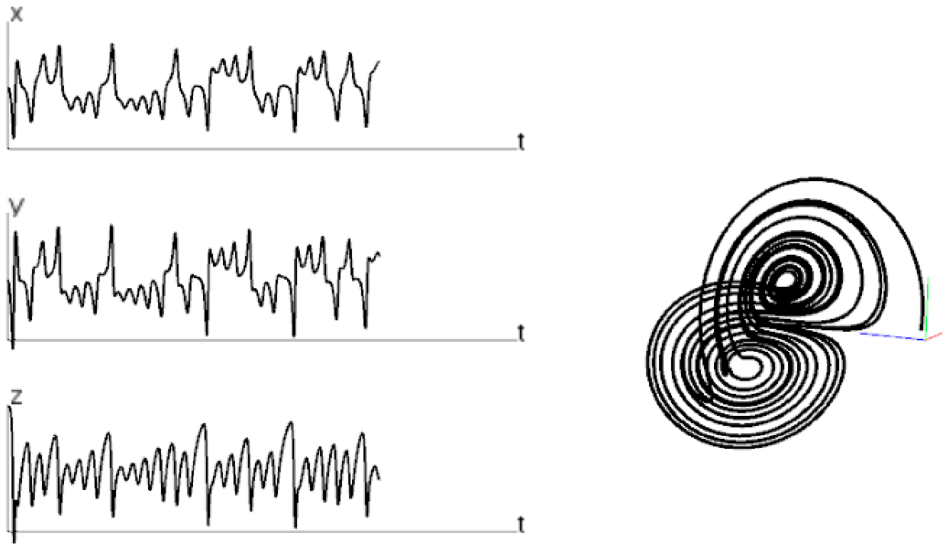


Figure 3.1: Trajectory of Lorenz Attractor in 3 Dimension Space

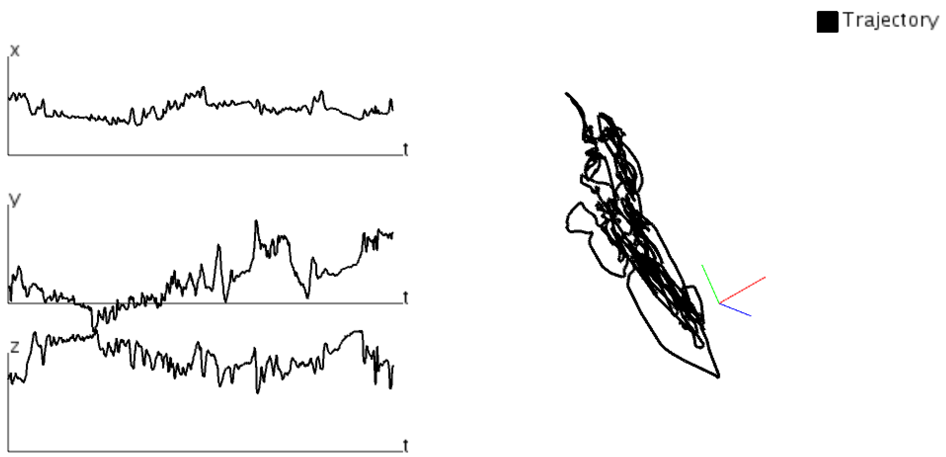


Figure 3.2: Trajectory of Head Movement in 3 Dimension Space

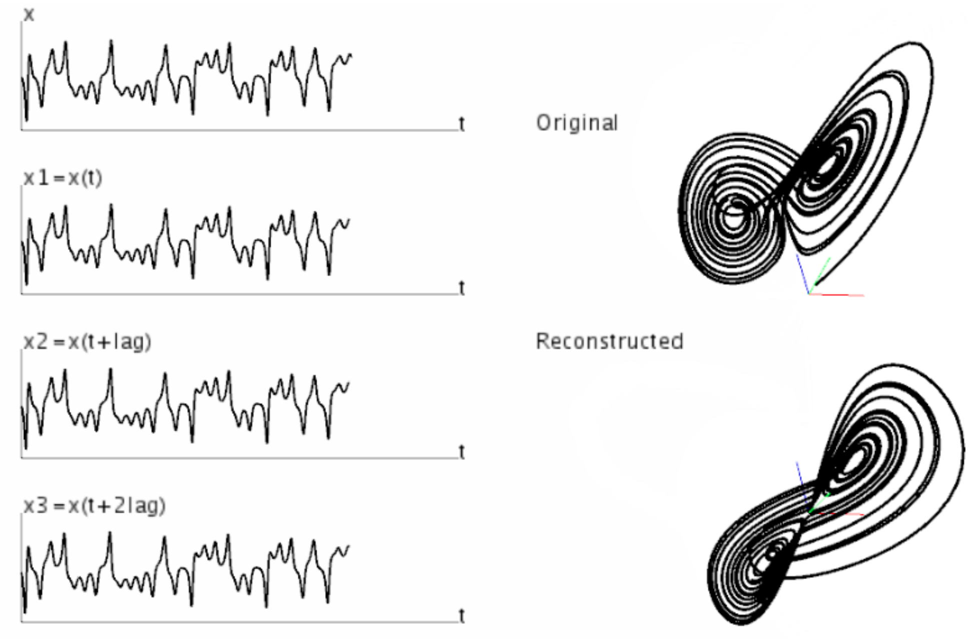


Figure 3.3: Reconstruct Lorenz Attractor in the 3 Dimension Phase Space

of a movement might not be able to be collected. For example, what if there is only the X-axis time series motion data of the Lorenz Attractor? In this case, a phase space reconstruct could be done to recover the original motion with a cost of fidelity in two steps:

i. Label all the N points of the time series T (X-axis, for example).

$$T_i = P_1, P_2, P_3, ..., P_N \tag{3.3}$$

ii. Reconstruct points V in the phase space with the embedding dimension m using the time-lagged embedding method with the delay τ.

$$V_i = P_i + P_{i+\tau} + ... + P_{i+(m-1)\tau} \tag{3.4}$$

Figure 3.3 shows the reconstructed time series from the original X-axis time series and the reconstructed Lorenz Attractor in the phase space, dim = 3, delay = 5. As we can see in the figure, the reconstructed Lorenz Attractor looks like a compacted original Lorenz Attractor. This illustrated the capability of the phase space reconstruction process in recovering the original system states. Figure 3.4 shows the reconstructed time series of head movement in the phase space, dim = 3 (3 dimension is chosen for demonstration purpose since most of us can only understand 3 dimension movement. Actually, the dimension can be higher than

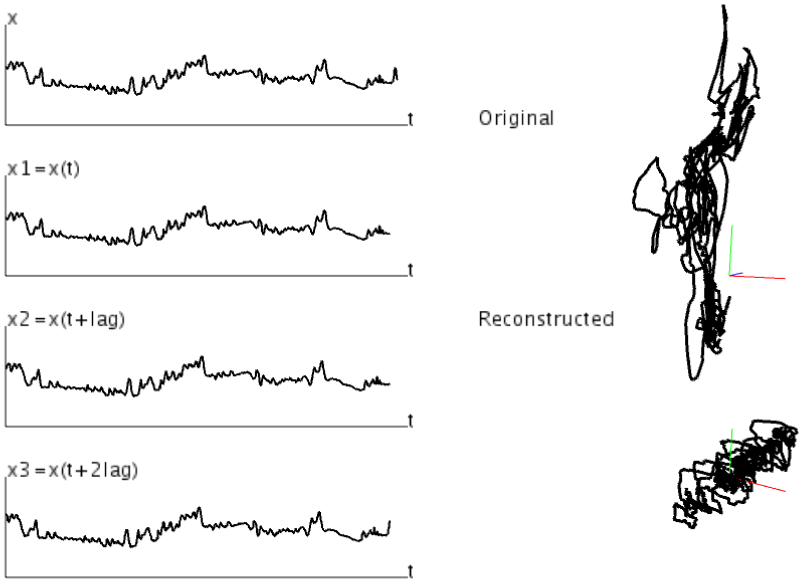


Figure 3.4: Reconstruct Head Movement in the 3 Dimension Phase Space

3.), delay = 15. The reconstructed head movement looks quite different from the original head movement. The trend in the other two dimension (head yaw and roll) is reduced. On the other hand, it illustrates the spiral movement in head nodding. The head will revisit certain places in the phase space over time, that is, the recurrence of head nodding.

Before performing the RQA, a distance matrix (DM) needs to be created by calculating the distance between every two points in the reconstructed points V. The distant D of two reconstructed points V_i and V_j can be calculated with the Euclidean distance:

$$D_{i,j} = \sqrt{\sum (V_i - V_j)^2} \tag{3.5}$$

When doing xRQA, the distance D is between the points of two different time series. The same process can be performed to get the reconstructed points W_j of the other time series. Then calculate the distance:

$$D_{i,j} = \sqrt{\sum (V_i - W_j)^2} \tag{3.6}$$

For every point P_i, P_j in time series T, a $D_{i,j}$ is calculated. Thus, a distance matrix DM is created. For every i, j in N, a matrix of RM can

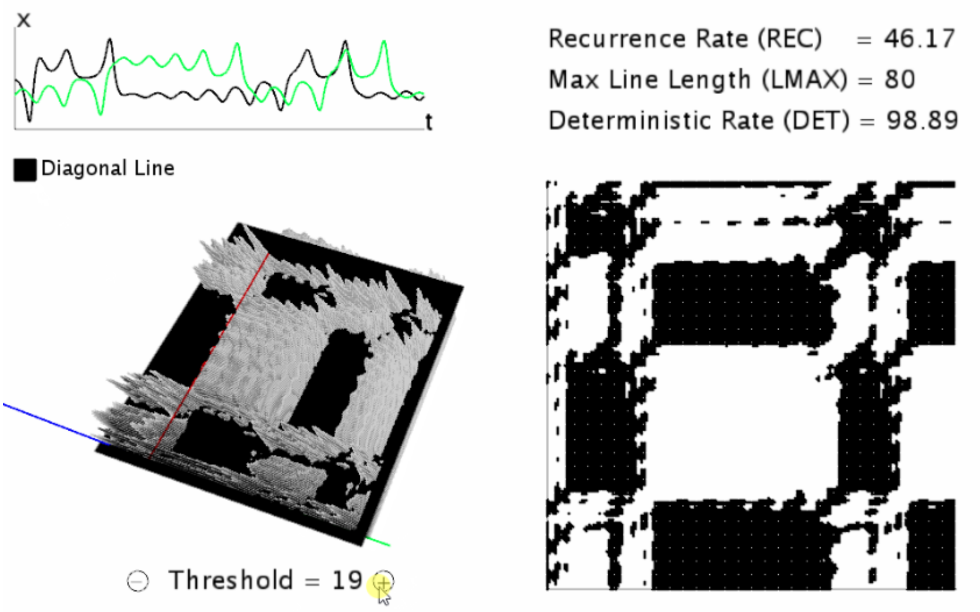


Figure 3.5: Creating Recurrent Plot with Thresholding Radius

be produced with:

$$RM_{i,j} = \begin{cases} 1, DM_{i,j} >= Radius, \\ 0, DM_{i,j} < Radius \end{cases} \tag{3.7}$$

With RM, the Recurrent Plot (RP) can be created. Simply draw a dot when $RM_{i,j}$=1 in the plot, and leave it empty when $RM_{i,j}$=0 (Figure 3.5). Figure 3.6 is an example of recurrent plot.

The recurrent plot can indicate how two motions repeat each other over time. However, it cannot represent how two groups of motions repeat each other over time. Romano et al. (2004) introduced Joint Recurrence Quantification Analysis for this problem. It can create RP for multiple groups by multiplying K RMs with the equation:

$$JRM = \prod_{k=1}^{K} RM_{i,j}^{k} \tag{3.8}$$

The JRP visualizes the simultaneous occurrence of recurrences. However, it diminishes the information presented as long as the processed RMs increase. Here, another simple method to produce RP for groups – the colored recurrent plot is introduced. The colored RP simply sum all

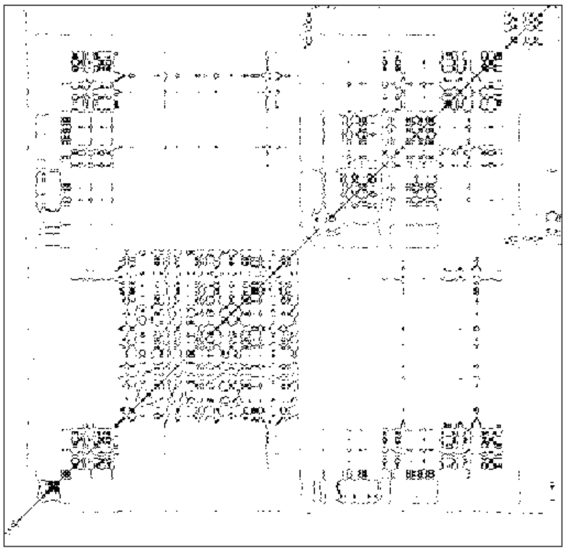


Figure 3.6: The Recurrent Plot of Head Nodding

the K RMs with the equation:

$$CRM = \sum_{k=1}^{K} RM_{i,j}^{k} \tag{3.9}$$

With the CRM, we can use the colored plot method to produce the colored recurrent plot (Figure 3.7).

The three most useful quantifications (%REC, %DET, LMAX) can be calculated with the equations:

$$\%REC = \frac{1}{N^2} \sum_{i,j=1}^{N} RM_{i,j} \tag{3.10}$$

%REC (recurrence): The percentage of recurrence points in the re-

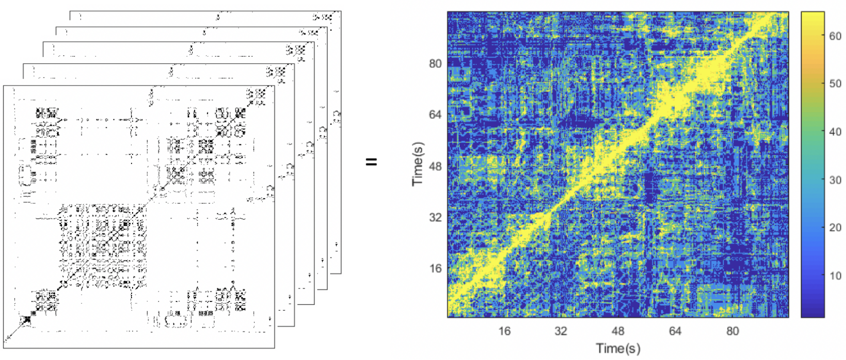


Figure 3.7: The Colored Recurrent Plot of Head Nodding

current plot.

$$\%DET = \frac{\sum_{l=l_{\min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \tag{3.11}$$

%DET (determinism): The percentage of recurrence points that form diagonal lines.

$P(l)$ is the histogram of the lengths l of the diagonal lines.

Diagonal line segments must have a minimum length defined by the line parameter. The name 'determinism' comes from repeating or deterministic patterns in the dynamic. Periodic signals (e.g., sine waves) will give very long diagonal lines. Chaotic signals (e.g., Hénon attractor) will give short diagonal lines. Stochastic signals (e.g., random numbers) will give no diagonal lines at all (unless the parameter RADIUS is set too high).

$$LMAX = \max\left(\{l_i;\ i = 1, \ldots, N_l\}\right) \tag{3.12}$$

LMAX (linemax): the length of the longest diagonal line segment in the recurrent plot, excluding the main diagonal line of identity (i = j).

LMAX is an important recurrence variable. The shorter the linemax, the more chaotic (less stable) the signal.

xRQA method is applied on the head nods time-series data. The three xRQA outputs (%REC, %DET, LMAX) as a whole indicate the coordination of head nods. A baseline chance coordination of the speaker and listener's nods is calculated by doing xRQA with randomly paired speaker's and listener's from the Natural condition. The head nods coordination in each condition is calculated for the speaker-listener pair.

3.4 Summary

In this chapter, the methodology approach for the research is described. The "Digital Chameleons" study is replicated with two customized experiments. The effectiveness of the virtual speaker would be evaluated with the same method as the original study and reassessed with the principal component analysis. The head nods would be analyzed with traditional linear statistical methods, frequency analysis method. The coordination of head nods would be evaluated with the nonlinear analysis method – Cross Recurrence Quantification Analysis (xRQA). Theprocess of xRQA is detailed elaborated.

Chapter 4

Experimental Immersive Virtual Environment Design

As noted in Chapter 3, a customized IVE is used to support human-agent interaction as well as human-human interaction. With the system, the "Digital Chameleons" paradigm is replicated in a somewhat more realistic virtual environment and with full-body avatars. The experimental environment is settled in the Human Interaction Lab of Queen Mary University of London. Participants take the first-person perspective on their avatar. The Vicon motion capture system is used to provide full-body motion capture. The design of the whole system is explained in the following sections.

4.1 System Architecture

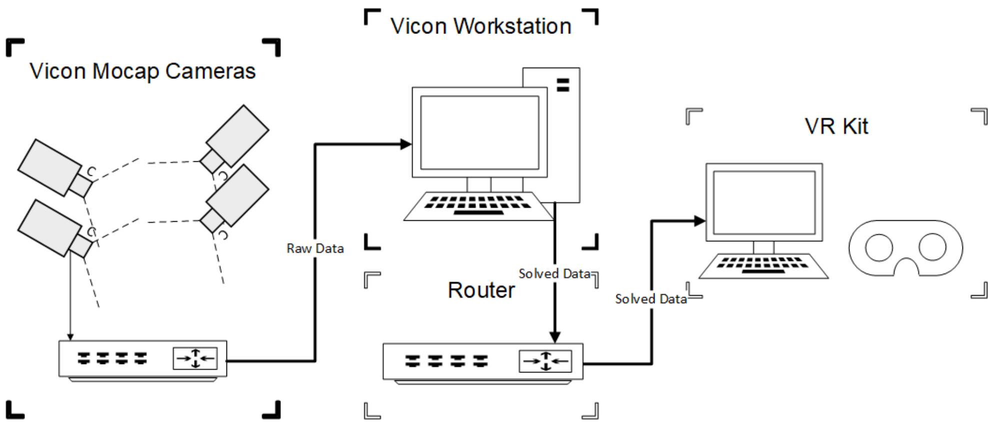


Figure 4.1: Physical Architecture of the System

The system integrates the full-body motion capture system and immersive virtual reality system. Particularly, Vicon Mocap system is used for natural nonverbal behavior input, and Oculus Rift HMD is used to present the virtual world. The Vicon Mocap system is quite advanced that can provide great precision in motion tracking and Oculus Rift HMD is able to present the best VR render quality. On the one hand, the equipment ensures the best immersive experience for the participants to create a great sense of body-ownership, thus, to provoke natural interaction in the virtual reality (Maselli & Slater, 2013). On the other hand, this IVE setting enables the great controllability of the experiment so that certain behavior of the participant can be manipulated.

Figure 4.1 shows the physical setting of the system. Motion data is first captured with the Vicon Mocap cameras, then streamed to the Vicon workstation for solving. The solved data is sent to the VR workstations and rendered with Unity3D and presented with HMDs. The equipment is listed in Table 4.1.

On the Vicon workstation, Vicon Blade takes the raw captured motion data – markers' positions in the space, generating skeletons with transition information including translations and rotations. The skeletons are used to drive virtual characters, termed solving the skeleton. The solved motion data is obtained with the Vicon DataStream SDKs. At the same time, Vicon Pegasus is used to do retargeting. Since Vicon Blade has its skeleton hierarchy, which is usually different from most humanoid virtual character's skeleton hierarchy, a retargeting process is necessary for the solved motion data to drive virtual characters correctly. The retargeted skeleton data can be accessed with Vicon Pegasus SDKs. The retargeted motion data is streamed to the VR workstation. One VR workstation is only able to support one HMD to present the virtual world. The VR workstation runs the Unity3D instance to produce the virtual environment.

4.2 Apparatus Settings

The Vicon motion capture system is settled in the Human Interaction Lab, Queen Mary University of London. Figure 4.2 is the schematic

Table 4.1: The Equipment Used in the Study

Item	Spec.	Unit
Vicon Mocap System	Optical Camera MX13 & MX40	12
	MX Link	2
	MX Net	1
	MX Control	1
	Host PC (Vicon Workstation)	1
	Vicon Blade (Software)	1
	Vicon Pegasus (Software)	1
VR System	Oculus Rift DK2 (HMD)	1
	Host PC (VR Workstation)	1
	Microphone	1
	Headphones	1
	Unity3D (Software)	1

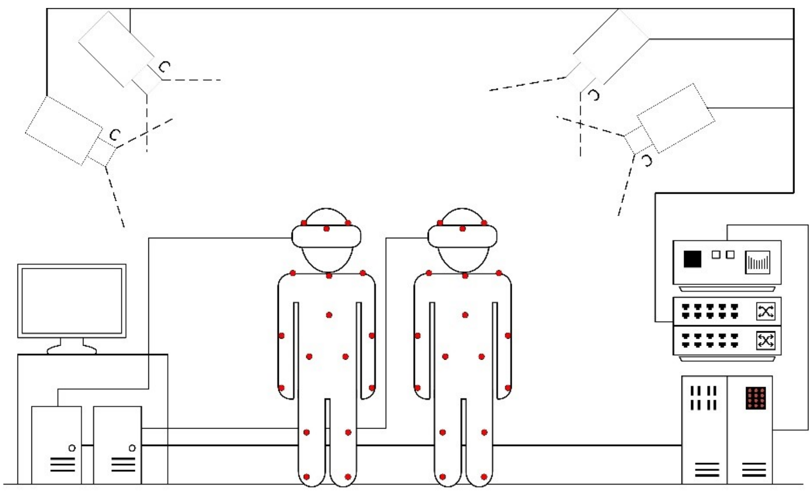


Figure 4.2: Schematic Diagram of the Capture Set-up in the Human Interaction Lab

Figure 4.3: Participant wore black suits attached with markers and Oculus Rift DK2 (HMD)

diagram of the capture set-up. 12 optical cameras are installed under the ceiling. These 12 optical cameras facing the middle of the room create a capture volume of about 2.5 (m) x 2.5 (m) x 2 (m). 2 middle-sized and 2 large-sized marker suits are prepared to fit different sized participants. The markers on the suits need to be carefully placed on the fixed certain positions for every participant (Figure 4.3). Participants need to wear lycra suits with markers attached and the Oculus Rift DK2 (HMD) in the standing position.

4.3 Virtual Environment

Since the effectiveness of the speakers with different head nods is investigated in the study, being behavioral realistic is far more important than being visual realistic. Participant's attention should be kept at the virtual speaker's head nods rather than the appearance of the virtual speaker or the virtual environment. Thus, in the design of the IVE, the following principle should be considered:

1. Use a blank IVE so that there would be nothing attractive in the IVE other than the speaker – this is how the original "Digital Chameleons" study was implemented. However, an empty space is unusual and different from a space that would facilitate social interaction.

2. Use the same environment as the laboratory in which the experiment conducted. This environment would also be not very attractive since it is the same as the real world. However, same as noted above, a laboratory is not like a space that would facilitate social interaction.

3. Use an IVE that correspond to the context of the scenario. Thus, it could help to facilitate the interaction. For example, our participants were all students. The task of our study was a virtual speaker delivering the student regulation about student ID card to the participants. Thus, an office environment could be used. The IVE being different from the real world would be a distraction from the virtual speaker. However, by giving participants free time to explore the IVE before the experiment, this distraction could be diminished.

Figure 4.4: The virtual environment in the participant's view

The third design is used in our study. The virtual environment is powered by Unity3D. It is designed as a virtual office in the daytime. There is furniture in the office such as desks, chairs, computers and some plants. Out of the office is a real street view of London. Figure 4.4 shows the virtual environment from the participant's point of view. Participants see their body movements in real-time. Participants also see a virtual character (a female character in this figure) stand in front of them.

Recently, Unity3D Game Engine as the Virtual Reality platform has become popular (Brookes, Warburton, Alghadier, Mon-Williams, & Mushtaq, 2019). Unity3D is a widely used 3-D game engine for developing video games, animations, and other 3-D applications, and it is growing in ubiquity. It is increasingly being used in research settings as a powerful way of creating 3-D environments for a range of applications (e.g., psychology experiments, surgical simulation, or rehabilitation systems). Unity3D has well-developed systems in place for vibrant graphics, realistic physics simulation, particles, animations, and more. Nevertheless, it does not contain any features specifically designed for the needs of human behavior researchers. an open-source software resource is set out to be produced to empower researchers to exploit the power of Unity3D for behavioral studies.

4.3.1 Virtual Character Library

There are also three different ways to design the virtual characters:

1. Use abstract virtual characters such as stick man, wooden mannequin (Latoschik et al., 2017), potato man (Gurion, Healey, & Hough, 2018). Abstract virtual characters could improve the agency (Argelaguet, 2016). On the other hand, they lost their social identity.

2. Use photo scanned models of the real speakers. This would be the ideal way for the study. However, it is time consuming and technically difficult. Furthermore, it would bring in a new uncontrollable variable – the appearance of the virtual speaker.

3. Use realistic characters which look not like the real speakers. This is a relatively cheaper solution than the second one. However, the inconsistent appearance of the virtual speaker and the real speaker might create some unknown effect. The good part is, using these characters, on the one hand, is good to facilitate social interaction, on the other hand, is good for the experimental control.

Overall, the third design is used. The virtual characters are created with Adobe Fuse. Adobe Fuse is free software for creating the virtual character. It provides us with a variety of 3D character components, including the face, hairstyle, clothes, and adjustable features such as height, arm length, etc. which made it very easy to customize our 3D characters. Figure 4.5 is the customize face interface of Adobe Fuse while creating a male character.

Four different virtual characters (2 males and 2 females) are created for the 4 conditions of the experiments, i.e., two interactants in mixed-gender. These virtual characters are bounded with skeletons automatically by the software. Therefore, they can be directly used in the Unity3D for animation as well as Vicon Pegasus for retargeting.

4.3.2 Animating Virtual Characters

Two techniques are used to animate virtual characters: recorded animation and real-time motion tracking. The recorded animation is usually used on intelligent virtual agents, while real-time motion tracking is generally used on avatars to create immersive experiences.

An IVE should synchronize virtual avatar's movements with its user's real-world movements. In our system, the Vicon Mocap system is used to track participants' actions. The motion data drives the virtual characters. The Vicon Pegasus plugin powers this function for Unity3D. Figure 4.6 shows a posture captured in Vicon Blade, retargeted in Vicon Pegasus, and animated in Unity3D. The Vicon Blade generates a general skeleton with movement information. This information is streamed to the Vicon Pegasus, which retargets the movement to a specific virtual character. Unity3D obtains the retargeted movement data with the Vicon Pegasus Plugin and renders the virtual character with the movement

Figure 4.5: Use Adobe Fuse to Create 3D Characters

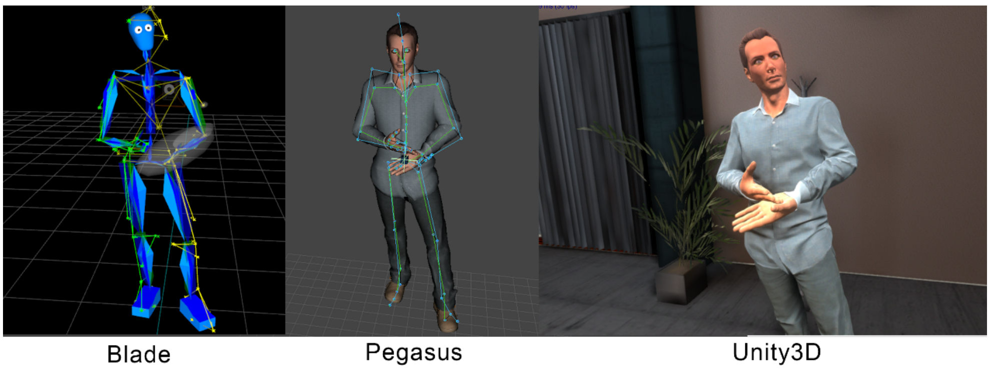


Figure 4.6: Captured Body Movement as Modeled in Different Stages of Re-targeting and Re-synthesis

data in the virtual environment.

The virtual character's facial expression is essential in the virtual non-verbal interaction. However, there is no equipment for real-time facial expression capture in the lab. Therefore, in this study, SALSA lip-sync-and-random-eyes is used to synthesize facial expression for the virtual character. SALSA lip-sync-and-random-eyes is a Unity3D plugin found in the Unity3D asset store to generate synthesized lip movements in accord with the loudness of the input voice and random eye movements and blinks for virtual characters. This technology cannot create facial expressions for virtual characters as well as facial expression capture techniques. Still, it is an easy, accessible and a similar technique as the one used in the "Digital Chameleons" study.

4.4 Motion Data Recording

As required for the research, all the motion data is recorded for every participant. Since the framerate will fluctuate when running the VR application, the motion data should not be recorded within each frame without a timestamp. The motion data can be captured in two ways: with a constant time interval (Method 1) or flexible time interval with a timestamp (Method 2). These two methods are compared in Table 4.2. Eventually, a constant interval data log (Method 1) is used as replay precise and multi-thread optimization are more critical in this study.

The data is recorded in the JSON format. It is structured as follows:

Table 4.2: Two Data Log Methods Comparison

Item	Method 1	Method 2
Timestamp	Optional	Compulsory
Time Interval	Constant	Flexible
Multi-thread optimizable	Yes	No
Storage usage	Constant	Flexible
Advantage	Replay Precise	Recording Precise

```
[
    {‘Head’: [x_0, y_0, z_0], ‘Neck’: [...], ...},
    {‘Head’: [x_1, y_1, z_1], ‘Neck’: [...], ...},
     ...
    {‘Head’: [x_t, y_t, z_t], ‘Neck’: [...], ...}
]
```

Where the motion data is regarded as a list of skeleton orientation objects, each item in the list (embraced with paired curly brackets and separated by comma) consists of a set of skeleton joints with their three-axis orientation information in degrees. 15 milliseconds is used as the recording interval for the motion data.

This motion data is stored as a file for each participant named with the participant's role and number, e.g., Speaker-01.json, Listener-01.json.

4.5 Manipulating Head Movement

In order to replicate the "Digital Chameleons" study, the IVE should be able to support manipulating the virtual character's head movement. In this study, three different manipulations are enabled with Unity3D scripts: i) head nods mimicry in 4s delay, ii) replay of previous head movements, and iii) no manipulation.

Different techniques are used to realize the manipulations. Figure 4.7 illustrates the pipeline for the head movement manipulations. First, the manipulated head movements need to be blended with the original body movements. The lateUpdate() function provided by Unity3D is used for the manipulated moves to override the original movements. Second, for the manipulation i, a delay is applied to the mimicry head movements. A timer is created to produce the delay. The delay is set to 4 seconds following the "Digital Chameleons" study. Meanwhile, a queue is used to buffer all the delayed head movements data. The enqueue and dequeue interval is a constant of 15ms (approximately 60Hz). Third, for the manipulation ii, all the movements data is logged at the beginning. With the previously logged movements file, the recorded head movements of the last people are applied to the virtual character. Also, the log interval is set to 15ms. Finally, for the manipulation iii, the head movements of participant are directly mapped to the virtual character..

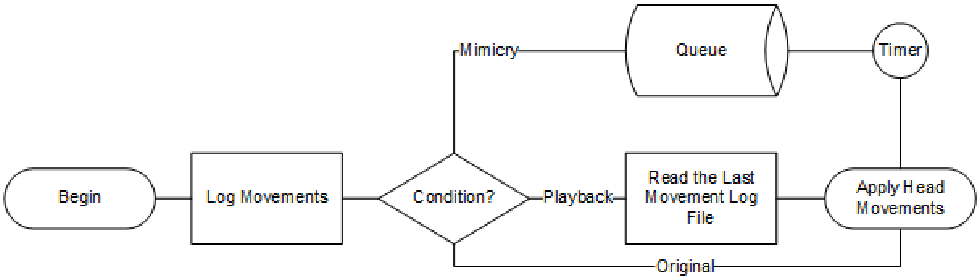


Figure 4.7: Pipeline of Head movement Manipulations

4.6 Multithreading

VR application requires high framerate for the optimal user experience. However, the avatar animating, motion data recording, and head movement manipulating processes could be I/O operation intense. These processes could block the render pipeline and decrease the display framerate. The multithreading technique is used to solve this problem. These processes are put in different threads apart from the main render pipeline. As a result, the render framerate of the VR application is kept above 80 Hz, which is the cap framerate of the Oculus Rift DK2 HMD.

4.7 Evaluating the System

In order to test the usability of the IVE, two pilot studies are run before the main study. One uses IVA as the speaker, the other one uses real participant embodied in an avatar as the speaker.

4.7.1 Supporting IVA Speaker-Listener Interaction

4.7.1.1 Procedure

Before the experiment, one female and one male confederates were recruited to produce the IVA's animation. They wore marker suits and performed a scripted speech about the college regulations relating to student ID cards (Appendix B). Their natural body movements were recorded by the motion capture system and used to create the IVA's animations.

In the experiment, the participants are asked to wear the marker suits. They are immersed in the IVE and embodied in a virtual avatar. In the IVE, a recorded message about student ID card is delivered by an

Figure 4.8: Head Movement of Participant in One Minute

IVA to the participant. After two minutes of interaction, the participant is asked to fill online questionnaires (Appendix C).

4.7.1.2 Participants

8 participants are recruited with email, posters, and flyers. 4 males and 4 females, age from 20 to 36 ($M = 29.5$, $SD = 2.08$). Each participant receives 7 pounds for their participation.

4.7.1.3 Results

The head movements of every participant are logged so that some preliminary analysis of the participants' behavior in the interaction could be done. Moreover, the open questions in the questionnaires helps a lot with understanding the usability of the IVE and whether participants have detected the mimicry or not.

Effectiveness of the agent Since there are only 8 participants, it is not valid to claim there is any significant effect of the agent. The survey data shows that the effectiveness of the agent varies between participants from a least of 1 (most disagree) to most of 7 (most agree). The mean is around 5, and Standard Deviation is about 1.5. The mean value for the effectiveness of the agent is relatively high.

Head Movement Figure 4.8 is the plot of one participant's head movements. The X-axis is time, and the Y-axis is the participant's head movement in degrees. There are 3 time series lines in different colors. Blue line for head nodding, red line for head shaking, and green line for head rolling. With head nodding, the positive value means pitching down, and the negative value means pitching up. With head shaking, the positive value means yawing right, and the negative value means yawing left. With head rolling, the positive value means rolling right, and the negative value means rolling left. Although the head movement of each participant was varied a lot, some patterns could still be identified in the plot. The participant had several different head behaviors in just one minute. The blue line in the red windows shows a nodding behavior. And the red and green line in the yellow window shows a head rolling and shaking. And the blue window shows a relatively stable head movement, indicating that the participant was looking straight and paying attention to the agent.

Qualitative Analysis of Questionnaires and Observation In the questionnaire, there are 4 open questions. Two questions are about interaction with the agent: (1) Please list any thoughts you may have about the interaction with the presenter; (2) Was there anything unusual about this interaction? Two questions are about the detection of agent movements: (1) Please write a few sentences about the presenter's LIP movements while speaking; (2) Please write a few sentences about the presenter's HEAD movements while speaking.

The answers from participants are sorted into a few themes; the number in brackets is the number of times participants had mentioned the theme:

Motion sickness (3): A few participants reported motion sickness from the virtual avatar. This motion sickness might come from the system latency and tracking problems resulted in misalignment between the embodied avatar and themselves.

Explore virtual environment (3): Some participants want to explore the virtual environment, and they started to look around and walk in the IVE. They were pulled back when they were going to hit the cameras.

Eye contact (5): One participant reported the agent was not looking at him/her. While another participant stated, the agent was staring at him/her. They said in a different attitude towards the eye contact: "(the agent) did not break eye contact", "(the agent had) good focus on me".

Imperfect animation (5): Some participants reported the weird hand of the agent. And body movements were strange when an agent walked in.

Too Intimidating (4): Some participants reported that the distance between them and the agent is too short, which made them felt too intimidating.

Abrupt (4): Some participants reported no real context of the interaction, and there was no natural interaction between them and agents (such as nodding, looking away, asking if understood).

Realistic (15): "The body movement created a realistic interaction". "The head moved according to the speech". "Head movement made realistically". "The pace is helpful". "Her manner made her seem approachable". "Friendly". "Natural". "Felt normal". "The head till to the side a little bit made it more natural".

Lip unmatched (9): Half of the participants said the lip movement was a bit out of sync. The other half think it was good.

Weird Head Movement (4): "The head dip randomly like she had fallen asleep". "The presenter seems like he is judging my look, made me uncomfortable". "Head moved too much".

From the result shows that most of the participants found the virtual environment and the virtual agent reasonably realistic. Most of the participants reported the realisticness comes from the body movement. No one detected the mimicry or playback head movement of the agent. Although it is not clear whether there is an effect of the mimicking agent with the data collected, some exciting thoughts of the participants are found. The result also pointed out the imperfecttion of the IVE. Quite a few participants reported for motion sickness, broken hand animation, and unmatched lip movement. This imperfection, on the other hand, reduced the realism of the IVE.

Apart from the questionnaires, some interesting phenomena is observed during the pilot study. The results show substantial individual variation between participants. Some participants would like to listen carefully to the people who were talking. In these cases, they would stare at the virtual agent and barely moved their head. Some other participants were very interested in the IVE. When they were immersed in the IVE, they would love to spend some time exploring, especially when their full-body was tracked. They would like to walk around or look around the IVE. In this case, their head movements increased a lot. Thus, there were two different attitudes in the questionnaire toward eye contact and head movement. Some participants said the agent did not break the eye contact and stared at him/her. This was because he/she was in a mimicry condition. If he/she listened very carefully to the agent and hardly moved his/her head, the agent did the same to him/her. This had the effect of making it appear that the agent was staring at him/her. Some other participants said the agent's head movement was weird because it sometimes dipped randomly like it had fallen asleep, or it was judging participant's appearance. This was also because the agent was in the mimicry condition and would reflect the participant's movement – the participant was looking up and down. The agent's reflection of his/her behavior at a delay causing misunderstandings. These observations highlight the fact that participants did not detect mimicry manipulation.

4.7.1.4 Discussion

The results prove that the IVE is able to replicate the "Digital Chameleons" study though a few limitations are identified through the pilot study.

The first limitation is the embodied experience. Because of the system latency, which might come from the network or graphic performance, participants would feel motion sick if they move around quickly. This could be improved by using better devices.

The second limitation is the participants' willing to explore the virtual world.This might be constrained because of the physical space limitation of the human interaction lab. In the future, the interior of the human interaction lab could be copied into the IVE. On the one hand, it could create a more significant realistic experience. On the other hand, participants would feel less curiosity about the virtual world so that their exploration intention is reduced and focus on the main task.. Moreover, let participants perform a short training or familiarization task can also help participants to focus on the main task and lead to better control of the experiment.

Improved Mocap system is needed to create better animation for an agent to deliver a persuasive message. And the lip movement could be improved by integrating face capture devices to HMD.

At this stage, the "Digital Chameleon" effect cannot be promised to emerge in the study. The mimicking agent might not be more persuasive than the agent with canned animation of a real speaker. The preliminary result suggests that the Digital Chameleon effect might not work well here.

4.7.1.5 Conclusion

In this pilot study, the IVE is tested by replicating the original "Digital Chameleons" study. The results suggest the IVE is able to support the replication of the 'Digital Chameleons' study. However, the system needed to be improved before the main study. The improvements could be made from the accuracy of the motion tracking system and the performance of the IVE. Meanwhile, preliminary data suggests that an IVA mimics its interactant is not necessarily more persuasive than an IVA plays the recorded animation of a speaker. And participants are not able to detect the mimicry IVA.

4.7.2 Supporting Avatar Speaker-Listener Interaction

4.7.2.1 Procedure

The second pilot study is conducted to test the usability of the system, where two participants are presented at the same time. Both of them need to wear the marker clothes. The Mocap system tracks both of the two participants. But only the listener wears the HMD and is immersed into the IVE.

Before the experiment, both participants are told to do a dialogue task in the IVE. One as the speaker, the other as the listener. The speaker participant is asked to deliver the message on a piece of paper. His/her full-body movements are tracked and rendered as the virtual speaker. The listener participant is immersed into the virtual office and embodied in the other virtual character. He/she is presented with the message of the regulation of student ID card by the virtual speaker. After the speaker finished the speech, the listener is asked to exit from the IVE and fill an online questionnaires (Appendix C). After the listener finished the online questionnaires, he/she is asked to leave. The study is using between-subjects design, each participant can only be the listener for once.

4.7.2.2 Participants

8 participants are recruited from a participant recruiting platform, age from 18-20 ($Mean = 18.75$, $SD = 0.707$), 3 Males and 5 Females. They receive 4 credits for participating in the study. The post-study survey showes that they are all from the same class and familiar with each other.

4.7.2.3 Results

The results are mainly obtained from the online questionnaires of the listeners right after the speakers finished their speech. Since the sample size is pretty small, only the descriptive data and qualitative analysis is presented here.

Effectiveness of the speaker The results show a pretty high rating on all the measures. The mean effectiveness of speaker is about 5.4 which is higher than the mean effectiveness of agent reported in Section 4.7.1. The Standard Deviation of effectiveness of speaker is also smaller than the SD of effectiveness of agent in Section 4.7.1.

Qualitative Analysis of Questionnaires The open questions which asked about the interaction, lip movements, and head movements of the virtual speaker are analyzed. Although the answers in the questionnaires do not havetoo many words, they are quite positive comments (Table 4.4).

Table 4.3: The Effectiveness of Speaker in Natural Condition.

Measures	Mean	SD
Agreement	6.22	0.86
Impression	6.49	0.829
Social Presence	4.57	0.376
Effectiveness	5.41	0.512

Table 4.4: The Answers for the Open Questions. The number in the bracket represents how many times different participants repeated the answer.

Question	Answers
Interaction	Easy, Nice(2), Helpful(2), Informative(3), Friendly
Lip movement	Clear, Fine, Normal(2), consistence(2), Wired
Head movement	Subtle(5), Unreal, Normal, Looked down, Good

4.7.2.4 Discussion

The comparison of the speaker's effectiveness between naturalcondition and recording condition (Section 4.7.1)indicates there might be some differences between human-human interaction and human-agent interaction. It seems that a listener would rate it higher in impression, social presence, and effectiveness when interacting with a virtual speaker embodied with a real speaker than interacting with an intelligent agent.

However, it is not clear that these effects were from the presence of the real speaker, the relationship between speaker and listener, the voice source from reality (not recording), or just the bad performance of the actors of which recorded movements animated the agent. This is probably because they are fimiliar with each other. This suggests that when paring participants, it would be better to make sure they are not familiar with each other.

The qualitative analysis indicates that the IVE system for the study is well designed. There are also some problems to be fixed. For example, 5 out of 8 participants reported the head movements of the virtual speaker were subtle. This might result from the failure of the behavior manipulation pipeline, or it was the real situation – their head did not moved much. Another participant reported the virtual speaker always looked down. This might because the real speaker had to read the message from a piece of paper with this pose. This could be solved by adding a virtual paper in front of the virtual speaker to provide the visual cue for the listener.

4.7.2.5 Conclusion

In the pilot study, the system is tested with 8 participants. They were running the experiment in pairs and being the listener and speaker in succession. The effectiveness of the speaker in the live condition and recording condition is compared. The result indicates that the speaker might be more effective in natural interaction than delivering a recorded animation in virtual reality. However, this effect might be because of some other confound variables which need further evidence. The reports from the participants proved the success of the system as well as the limitation of the system. Further improvement needs to be done to run the main study.

4.8 Summary

This chapter described the design of the customized IVE, including the architecture of the system, the lab setting, the virtual environment, and virtual character design, the avatar animating, motion recording, head movement manipulation pipelines. The system is evaluated with two pilot studies. The results proved the usability of the system.

Chapter 5

Evaluating the "Digital Chameleons" Paradigm

In this chapter, two empirical experiments utilized the customized IVE are presented. The first study replicates Bailenson and Yee's setting, which uses an IVA to perform as the speaker. In the second study, real speakers and a relatively natural setting is used. These experiments try to replicate the "Digital Chameleons" study. The results are detailed explained in the following sections.

5.1 IVA Speaker and Listener Interaction

The pipeline described in Chapter 4 enables us to replicate the "Digital Chameleons" paradigm. The replication includes some small changes to the materials and procedure of the original study.

Overall, each participant is immersed in an IVE and listen to a virtual agent read the persuasive message. The agent will mimic the participant's head movement or do a playback the head movements from the previous participant or play a prerecorded animation from an actor who was reading the message.

In this section, the procedure of the study is introduced first; then the measures, results of the study is explained; and finally, the results is discussed.

5.1.1 Participants

52 participants were recruited by email, posters, and through a participant panel. Each participant received 10 pounds or 4 credits for their participation. One participant is excluded because of problems in data logging. The final sample consists of 29 female and 22 male students between 18 to 52 (M=21.75, SD=5.836). None of the participants report severe motor, auditive or visual disabilities/disorders.

5.1.2 Procedure

The experiment is running in the Human Interaction Lab, Queen Mary University of London. Participants come to the lab. They are introduced to the purpose of the study:

> This is one of a series of studies to understand how we could run student information service effectively within the immersive virtual environment (IVE) or virtual reality (VR). The whole process will take about 60 minutes.

and asked to sign a consent form (Appendix D).

Participants are instructed to wear marker suits and Oculus Rift DK2 HMD. After putting on the capture suits and the HMD, they are given time for free exploration of the IVE (Figure 5.1). In the IVE, participants are standing in an office. They are welcomed by an IVA standing next

Figure 5.1: A Participant Is Exploring the IVE and Motion Tracking

to the desk and receive a short introduction as a training session. In the training session, a short message (around 30 seconds) about QMUL regulation on student course attendance (see Appendix A) is delivered by the IVA. It will perform the prerecorded actions from a confederate. The animation is a speech about the attendance regulation. Its movements are recorded with the Vicon Mocap system and voice is recorded with a microphone. Besides, its facial expression is automatically generated with SALSA lip sync and random eyes. The training session is provided to help participants to get familiar with the virtual environment and to understand the experiment procedure.

Figure 5.2: The IVA Is Delivering the Message to A Participant

Table 5.1: Number of Participants in Each Condition by Experiment Trial

Experiment Trial	Mimic	Playback	Recording	Total
1	18	16	17	51
2	17	16	18	51
Total	35	32	35	102

Next, a two minutes message is delivered by the IVA(Figure 5.2) asking students always to carry their student ID card (see Appendix B).

After the agent finished their speech, the participants are asked to take off the HMD and fill out an online questionnaire (see Appendix C).

Once the participants completed the questionnaire, they reenter the IVE and repeat the procedure in a different condition. The gender of the agent is switched, and itsbehavior is randomly assigned to a different experimental condition. Participants are asked to fill the questionnaire again after the experiment.

Participants are assigned to the mimicry, playback, and recording conditions in random order:

1. Mimic – In the mimic condition, the agent's head nods exactly mimic those of the participant at a 4-s delay.

2. Playback – In the playback condition, the agent's head nods are an exact replay of the nods from the previous participant. This ensured that the agent move as much in the mimic and playback conditions but with different timing.

3. Recording – In the recording condition, the agent play a captured movement of an experiment confederate delivering the message.

The experiment is conducted with a mixed of between/within-subjects design. Every participant takes part in the experiment twice. In this setting, every participant hears the same message twice. There are 35 participants each in the mimic and recording condition, 32 participants in the playback condition (Table 5.1).

In all the three conditions, only the head pitch of the agent will be manipulated, and the yaw and roll of the head are kept as the recording condition.

Table 5.2: The Components of the Agent's Effectiveness Measures

Measure	# items	α	Condition	Mean	SD
Agreement	4	0.891	Mimic	5.37	1.08
			Playback	5.55	1.04
			Recording	5.59	1.12
Impression	12	0.923	Mimic	4.56	1.12
			Playback	5.06	1.1
			Recording	4.71	1.0
Social Presence	7	0.683	Mimic	3.88	1.00
			Playback	4.28	0.86
			Recording	3.7	0.9
Effectiveness	23	0.917	Mimic	4.50	0.91
			Playback	4.91	0.89
			Recording	4.56	0.83

5.1.3 Effectiveness of the IVA Speaker

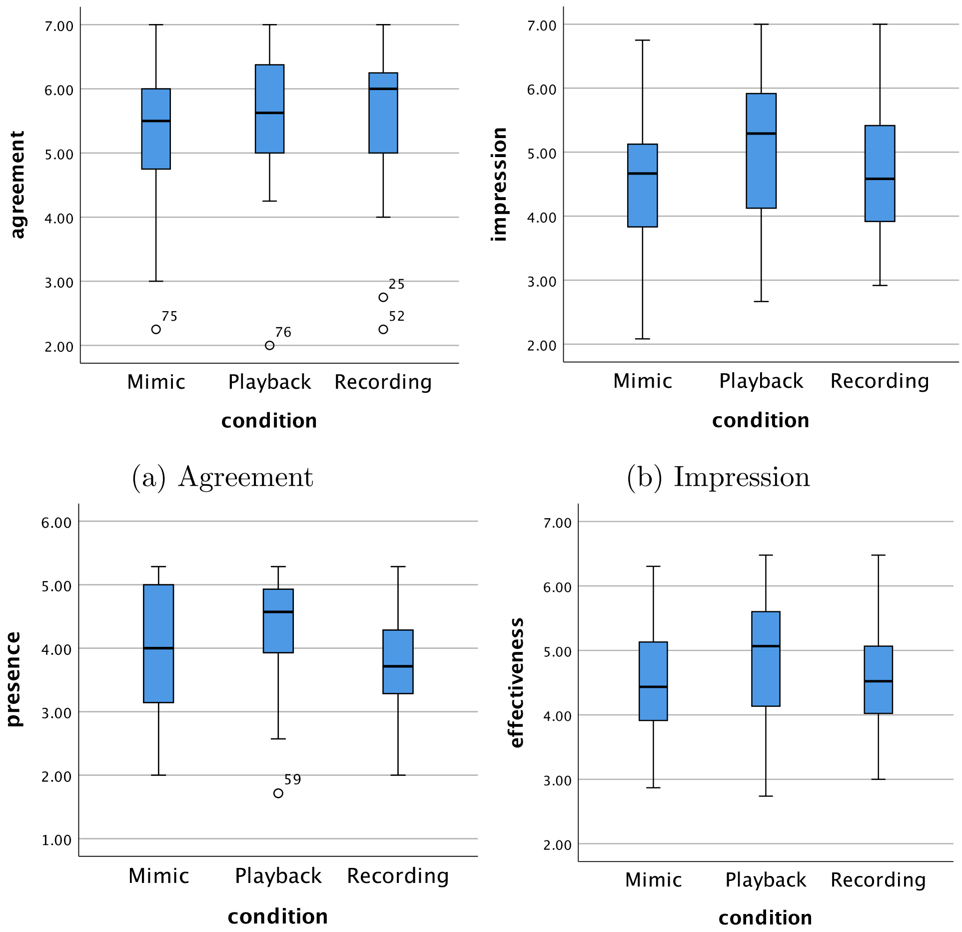
The composite measure of the agent's effectiveness from the "Digital Chameleons" study is replicated by taking the mean response to the 4 agreement questions (how much the participant agreed with the agent's persuasive message), 12 questions on impressions of the agent, and 7 questions on the agent's social presence. To provide a conservative test of the "Digital Chameleons" effect a non-directional null hypothesis is used:

Null Hypothesis 1 *There is no difference in the effectiveness of the agent between the different conditions (mimic or playback or recording).*

5.1.4 Results

5.1.4.1 Original Composition

The effectiveness of the agent is a composite measure of 23 items. The Cronbach's alpha for this is 0.917. A Shapiro-Wilk test shows that ratings of effectiveness does not deviate significantly from a normal distribution (p=0.482). Table 5.2 shows the descriptive statistics of the components of the agent's effectiveness measure. While the mean ratings of agreement, impression, effectiveness of agents are above the medium of 4 in all the 3 conditions, the mean ratings of agent's social presence are below medium in the mimic and recording conditions, above medium in the playback



(a) Agreement (b) Impression

(c) Social Presence (d) Effectiveness

Figure 5.3: The Boxplots of the Components of the Agent's Effectiveness Measures

condition. Figure 5.3 is the boxplot of the components of the agent's effectiveness measures.

As the experiment used a mixed between/within-subjects design, a Generalized Linear Mixed Model (GLMM) analysis is conducted. In the GLMM analysis, participant's order is put in the repeated measure; agent's agreement, impression, social presence, effectiveness are the targets; agent's head nodding condition, participant's order are the fixed effects; subject is included as random effects. Table 5.3 shows the Bayesian Information Criterion (BIC) for model selection for the agent's effectiveness measures. The table suggests the model using a log normal link function, with no fixed effects interaction and included random intercept has the least BIC of -54.429. This model is selected for the analysis of the agent's effectiveness measures.

Table 5.3: Model Comparison of GLMM for the Agent's Effectiveness Measures

Distribution & Link	Interaction	Random Intercept	BIC
Normal Identity	No	No	316.793
Normal Identity	Yes	No	315.329
Normal Identity	Yes	Yes	266.149
Normal Log	No	No	-17.497
Normal Log	Yes	No	-12.104
Normal Log	No	Yes	-54.429

Table 5.4: Test of Fixed Effects for the Agent's Effectiveness Measures

Target	Source	F	df1	df2	Sig.
Agreement	Corrected Model	0.120	3	98	0.948
	Condition	0.157	2	98	0.855
	Order	0.055	1	98	0.816
Impression	Corrected Model	1.256	3	98	0.294
	Condition	1.857	2	98	0.162
	Order	0.106	1	98	0.746
Social Presence	Corrected Model	4.072	3	98	0.009
	Condition	3.517	2	98	0.033
	Order	3.975	1	98	0.049
Effectiveness	Corrected Model	1.174	3	98	0.324
	Condition	1.66	2	98	0.195
	Order	0.111	1	98	0.740

Table 5.5: Estimates of Fixed Coefficients for the Agent's Social Presence

Item	Coefficient	Std. Error	t	Sig.	95% Conf. Int.	
					Lower	Upper
Intercept	1.29	0.039	33.055	<0.001	1.213	1.368
Cond.=1	0.071	0.03	2.337	0.019	0.012	0.131
Cond.=2	0.065	0.029	2.252	0.027	0.008	0.122
Cond.=3	0	.	.	.	.	.
Order=1	0.042	0.021	1.994	0.049	0.00	0.083
Order=2	0	.	.	.	.	.

Table 5.6: Estimates of Covariance for the Agent's Social Presence

Item	Estimate	Std. Error	Z	Sig.	95% Conf. Int.	
					Lower	Upper
Residual(Order=1)	0.15	0.073	2.066	0.039	0.058	0.387
Residual(Order=2)	0.208	0.074	2.828	0.005	0.104	0.416
Random(Intercept)	0.05	0.012	4.029	<0.001	0.031	0.081

Table 5.7: Pairwise Comparison of Agent's Social Presence by Condition

Condition Pair	Contrast	Std. Error	t	df	Adj. Sig.
Mimic - Playback	0.021	0.115	0.181	96	0.856
Mimic - Recording	0.286	0.111	2.582	96	0.011
Playback - Recording	0.265	0.115	2.301	96	0.024

Table 5.8: Pairwise Comparison of Agent's Social Presence by Order

Trial Pair	Contrast	Std. Error	t	df	Adj. Sig.
First - Second	0.186	0.082	2.274	96	0.025

Table 5.4 shows the fixed effects in GLMM on the agent's effectiveness measures. The results show no model can be built to predict agent's agreement, impression and effectiveness by condition or order; no significant fixed effect of agent's agreement, impression and effectiveness on condition. However, a GLMM can be built to predict the social presence of an IVA with condition and order (p=0.018). There is significant fixed effects of agent's social presence on condition and order. Table 5.5 shows the estimates of the fixed coefficients for the agent's social presence. Table 5.6 shows the estimates of covariance parameters for the agent's social presence. There are variability of different order on agent's social presence. There is a significant random intercept effect. This suggests that there are unmeasured variables interactions for which the agent's social presence appears random.

Pairwise comparisons of agent's social presence on condition and order are shown in the Table 5.7 and Table 5.8. The agent's social presence is higher in mimic (Mean Diff.=0.286, p=0.011) and playback (Mean Diff.=0.265, p=0.024) conditions than in the recording condition. The agent's social presence is higher in the first experiment trial than in the second trial (Mean Diff.=0.186, p=0.025).

5.1.4.2 Composition with Principal Component Analysis

The principal component analysis on the 23 items reveals five factors that together accounted for 73.8% of the total variance. All factors have eigenvalues of over 1. The item is considered to be valid if its absolution factor loading is greater than 0.3. The factor loadings of each item and the reliability of each factor are shown in Table 5.9. The first factor can be described as effectiveness because it included all the items. The second factor can be described as agreement because it included

Table 5.9: Factor Components and Loadings in IVA-Listener Interaction.

Factor	α	Eigenvalue	Item	Factor Loading
Effectiveness (23 Items)	0.917	10.2	Agreement	0.48
			Valuable	0.60
			Workable	0.52
			Needed	0.43
			Friendly	0.75
			Likeable	0.81
			Honest	0.59
			Competent	0.59
			Warm	0.75
			Informed	0.56
			Credible	0.58
			Modest	0.59
			Approachable	0.77
			Interesting	0.71
			Trustworthy	0.75
			Sincere	0.77
			Enjoyed	0.76
			Current Situation	0.83
			Isolated	-0.31
			Meet Again	0.73
			Comfortable	0.77
			Cooperative	0.78
			Embarrassed	-0.38
Agreement (7 Items)	0.805	2.5	Agreement	0.72
			Valuable	0.62
			Workable	0.65
			Needed	0.75
			Warm	-0.33
			Modest	-0.38
			Approachable	-0.36
Authority (8 Items)	0.869	1.9	Friendly	-0.34
			Honest	0.55
			Informed	0.47
			Credible	0.41
			Modest	0.48
			Current Situation	-0.30
			Meet Again	-0.47
			Comfortable	-0.31
Credible (3 Items)	0.310	1.3	Informed	0.44
			Credible	0.48
			Isolated	0.44
Isolated (2 Items)	0.316	1.1	Isolated	0.57
			Embarrassed	0.48

Table 5.10: The Descriptive Statistics of The Main Factors.

Factor	Condition	Mean	SD
Effectiveness	Mimic	-0.17	1.03
	Playback	0.28	1.01
	Recording	-0.08	0.93
Agreement	Mimic	-0.01	1.05
	Playback	-0.17	0.94
	Recording	0.16	0.10
Authority	Mimic	-0.25	1.0
	Playback	-0.15	0.76
	Recording	0.39	1.08

Table 5.11: Model Comparison of GLMM for the Main Factors

Distribution & Link	Interaction	Random Intercept	BIC
Normal Identity	No	No	291.715
Normal Identity	Yes	No	288.933
Normal Identity	Yes	Yes	271.2
Normal Log	Yes	Yes	525.638

4 agreement items, modest and approachable. The third factor can be described as authority because it included honest, informed, credible, modest as positive loadings, and friendly, current situation, meet again, comfortable as negative loadings. The fourth factor can be described as credible because it included informed, credible, and isolated. The last factor can be described as isolated because it included isolated and embarrassed. The scores for each factor are calculated for each participant using a regression method.

The reliability test suggests that the last two factors are invalid at reliability performance ($\alpha < 0.7$). Table 5.10 shows the descriptive statistics of the three main factors. A Generalized Linear Mixed Model (GLMM) analysis is conducted. In the GLMM analysis, participant's order is put in the repeated measure; agent's effectiveness, agreement and authority are the targets; agent's head nodding condition, participant's order are the fixed effects; subject is included as random effects. Table 5.11 shows the Bayesian Information Criterion (BIC) for model selection for the main factors measures. The table suggests the model using a identity normal link function, with fixed effects interaction and random intercept included has the least BIC of 271.2. This model is selected for the analysis of the main factors measures.

Table 5.12 shows the fixed effects in GLMM on the three main effects.

Table 5.12: Fixed Effects in GLMM for the Main Factors

Target	Source	F	df1	df2	Sig.
Effectiveness	Corrected Model	0.582	5	93	0.714
	Condition	1.103	2	93	0.336
	Order	0.0	1	93	0.984
	Condition*Order	0.272	2	93	0.763
Agreement	Corrected Model	1.533	5	93	0.187
	Condition	2.75	2	93	0.069
	Order	0.005	1	93	0.994
	Condition*Order	0.609	2	93	0.546
Authority	Corrected Model	1.984	5	93	0.088
	Condition	3.953	2	93	0.023
	Order	0.467	1	93	0.496
	Condition*Order	0.41	2	93	0.665

Table 5.13: Estimates of Fixed Coefficients for Authority

Item	Coefficient	Std. Error	t	Sig.	95% Conf. Int.	
					Lower	Upper
Intercept	0.304	0.197	1.542	0.127	-0.088	0.696
Cond.=1	-0.342	0.27	-2.537	0.013	-1.22	-0.149
Cond.=2	0.065	0.269	-1.27	0.207	-0.876	0.193
Cond.=3	0	.	.	.	.	.
Order=1	0.068	0.302	-0.226	0.822	-0.667	0.531
Order=2	0	.	.	.	.	.
Order=1* Cond.=1	0.401	0.471	0.852	0.397	-0.534	.337
Order=2* Cond.=1	0	.	.	.	.	.
Order=1* Cond.=2	0.075	0.479	0.157	0.875	-0.877	1.027
Order=2* Cond.=2	0	.	.	.	.	.
Order=1* Cond.=3	0	.	.	.	.	.
Order=2* Cond.=3	0	.	.	.	.	.

Table 5.14: Estimates of Covariance for Authority

Item	Estimate	Std. Error	Z	Sig.	95% Conf. Int.	
					Lower	Upper
Residual(Order=1)	0.567	0.171	3.321	0.001	0.314	1.023
Residual(Order=2)	0.276	0.134	2.052	0.04	0.106	0.717
Random(Intercept)	0.559	0.167	3.343	0.001	0.311	1.004

Table 5.15: Pairwise Comparison of Agent's Authority by Condition

Condition Pair	Contrast	Std. Error	t	df	Adj. Sig.
Mimic - Playback	-0.18	0.181	-0.993	93	0.323
Recording - Mimic	0.484	0.173	2.789	93	0.006
Playback - Recording	-0.304	0.183	-1.666	93	0.099

No significant difference is found on effectiveness or agreement between conditions or participant's order. However, there is a significant difference in authority between conditions ($F_{2,93} = 3.953$, p=0.023). Table 5.13 shows the estimates of the fixed coefficients for authority. Table 5.14 shows the estimates of covariance parameters for authority. There are variability of different order on authority. There is a significant random intercept effect. This suggests that there are unmeasured variables interactions for which authority appears random.

Pairwise comparison between conditions suggests that authority is higher in the recording condition than in the mimic condition ($t_{93} = 2.789$, p=0.006) (Table 5.15). No significant difference is found on authority by participant's order. No significant interaction effect between condition by order is found for all the factors as well. Figure 5.4 is the error bar chart comparing the authority between conditions. The error bar represents the 95% confidence of the mean authority.

5.1.5 Discussion

The results suggest that the differences in agent behavior have no significant effect on perceived persuasiveness and, as such, fail to replicate the basic "Digital Chameleons" effect. The results also show that the agent mimicking participant's head nodding has significant higher social presence than the nonmimicking agent. This suggests participants enjoyed more when interacting with a mimicking agent. However, this does not improve the overall agent effectiveness. The Null Hypothesis 1 cannot be rejected. Two possible explanations for this are: i) the "Digital Chameleons" effect is not reliable across different experimental situations; ii) this experiment fails to replicate some critical features of the original.

The results broadly favor the first explanation, partly because it highlights some critical problems with the automatic mimicry idea and its implementation in the "Digital Chameleons" study. First, a speaker

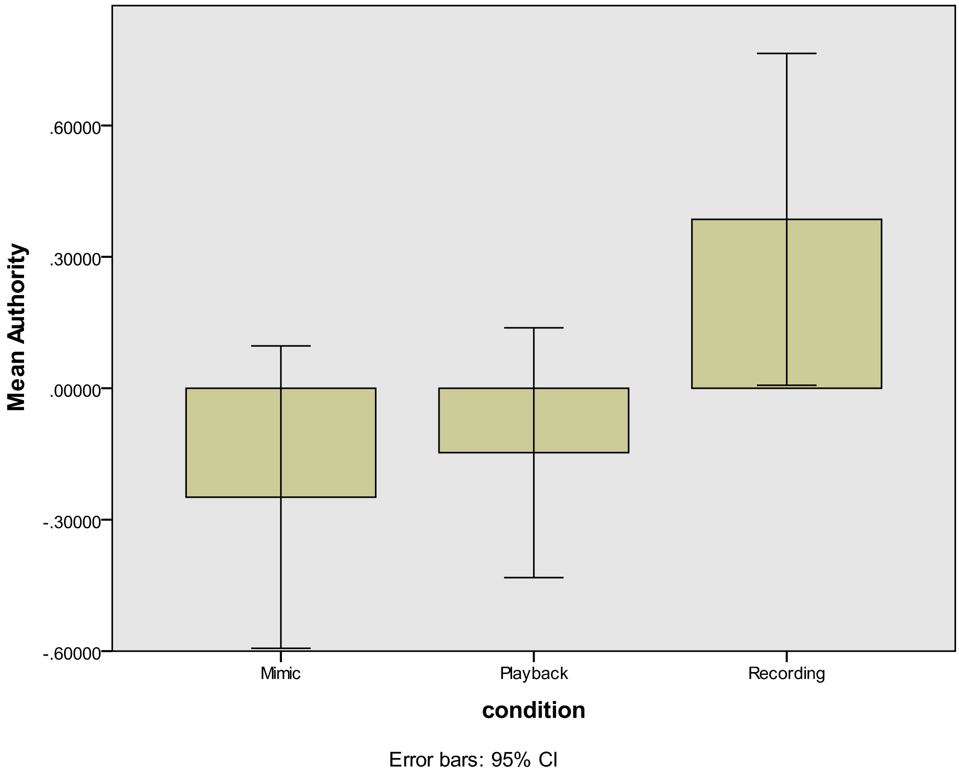


Figure 5.4: The Error-Bar Chart for Authority in Three Conditions.

usually moves their heads more frequently and in a wider variety of ways than listeners. If the speaker simply copies a listener's head movements, this represents a significant departure from normal behavior. This might have effects on persuasiveness in certain circumstances, but it breaks the balance of initiative typical of ordinary conversation. Secondly, if virtual agents 'blindly' mimic the participant's head movements, this breaks the relationship between their head movements and the content of their speech. It seems unlikely that this can improve persuasiveness unless, by chance, it lines up with what is being said. Third, the "Digital Chameleons" is based on the assumption that the chameleon effect could be produced by an automatic algorithm implemented on the virtual agent. However, this algorithm does not reproduce real human mimicking behavior, which is more flexible. Research shows that contingency might be more important than similarity in human mimicry behavior (Catmur & Heyes, 2013). Although unconscious human mimicry of an interaction partner might increase social influence, it does not entail that an agent would also gain social influence by implementing simple automatic mimicry rules. This highlights a gap in the theory which assumes we don't automatically mimic *everything* - since conversation would grind to a complete halt. However, it is unclear –from the point of view of implementation– how it is moderated in terms of timing, choice of behaviors, or with respect to particular social goals.

It is possible that an agent would gain social influence by mimicry if it applied a more selective algorithm to judge when it should perform a mimicry task in the social interaction context. Lee et al. (2010) compared the perception of inappropriate head nods generated by three different methods for a virtual agent: by a machine learning data-driven approach, by a handcrafted rule-based approach and by a human. The results suggest a data-driven approach has the best performance over all the three methods. While machine learning approach is very promising, there is lack of focus in modelling head nodding with this method. Perhaps the machine learning approach which generates head movements according to speech content and interaction could be explored in the future.

Nonetheless, the failure to replicate may be due to differences in the experimental setting. this study used a more sophisticated virtual environment and full-body interaction. The justification for this is greater naturalism, but it may also bring in some unpredictable factors such as the uncanny valley effect. Agents in the mimic and playback condition might be perceived as weird to the participants, given greater expectations about behavior or simply the greater range of behaviors. This might weaken or dilute the "Digital Chameleons" effect. Like Verberne et al. (2013) this effect might only work with certain behaviors, or it might need a long time to take effect. However, this line of argument leads to the conclusion that the effect is not robust. Another possibility is that students' attitudes toward the importance of ID cards have changed. Notably, the subjective ratings in the survey were relatively high, and this may obscure potential differences in persuasiveness.

Another concern is whether our participants actually perceived the differences of head-nodding behavior between the three conditions at all? There would be a high chance to get a null result if our participants did not notice the differences in head-nodding behavior between the three conditions. The factor analysis results rejected this hypothesis. The differences in agent behavior have no significant effect on perceived agent persuasiveness (effectiveness/agreement). This corresponds to the results noted above. However, the significantly higher recognized authority in the recording condition than in the mimic condition suggests that participants can perceive the differences of the head nodding behavior between the mimic condition and the recording condtion. The head-nodding behavior in the recording condition is significantly different from those in the mimic. And head-nodding behavior in the playback condition is not significantly different from the other two conditions. The results suggest that in the original study, Bailenson and Yee might be very much wrong with the manipulation of the agent's head nodding behaviors – mimic or playback of listener's head nods does not seem to differ from each other.

5.1.6 Conclusion

The results of this study provide no evidence of differences in participant's subjective estimates of a) social presence b) agreement with the agent c) general impression of the agent in the three different conditions. The effectiveness of the IVAs does not differ significantly across conditions where the IVAs are mimicking the participant's head nods or not. Furthermore, the PCA suggests that the IVA in the recording condition has higher authority than the IVA in the other two conditions. And the authority of the IVA in the mimic condition and playback condition does not differ from each other. This suggests 1) participants did perceive the different head nods across conditions 2) the head nods in the mimic condition and playback condition might not be significantly different. It might be wrong to compare the effectiveness of the agent mimicking the participant's head nodding and the agent using the previous participant's head nodding in the original "Digital Chameleons" study.

5.2 Avatar Speaker and Listener Interaction

Using the same virtual environment and motion capture system, the second experiment is conducted with a real speaker and listener. For most of the experiment trials, two participants are presented at the same time, one as the speaker and the other one as the listener. The experiment is repeated using this setting with the head-nodding manipulations the same as our first experiment.

In this section, the procedure is first introduced, then the arrangement of the participants is explained; and finally, the measures and results are discussed. In this study, slightly different results on persuasiveness is presented in the mimic condition. However, the effect is weak given the number of participants .

5.2.1 Participants

54 participants were recruited by email, posters, and through a participant panel. Each participant received 10 pounds for their participation. The final sample consists 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.

5.2.2 Procedure

The experiment is running in the Human Interaction Lab, Queen Mary University of London. Participants come to the lab. They are introduced to the purpose of the study:

> We are researching how do speakers use their nonverbal body language to persuade their listeners. The experiment consists of two parts. You will be the listener in the first part, and speaker in the second part. Each part will take around 30 minutes and followed by a 15 minutes gap for rest.

and asked to sign a consent form (Appendix E).

Participants are instructed to wear marker suits. The listeners wear the HMD. They will see the speaker in the IVE with a piece of virtual paper. The speakers, however, will not wear the HMD to read the material from a piece of A4 paper and see the real body movement of the listener. Figure 5.5 illustrates the situation in which two participants are doing the Experiment.

Before the experiment starts, the speaker would be told to try to act appropriately to gain more agreement from the listener. In the experiment, the speaker read out a 2 minutes long material (the same one in the first experiment - Appendix B) from a piece of A4 paper provided. Both speaker and listener's full-body movements are mapped to the corresponded virtual character. The listener will be embodied in a virtual character and see the speaker (of its virtual representer) delivering the message, through the HMD (Figure 5.6). After the session, the listener will take off the HMD and be asked to fill the online questionnaires (the same one in the first experiment – Appendix C).

In the experiment, the virtual speaker's head nodding is manipulated according to a randomly assigned condition:

1. Mimic – the virtual speaker's head nods exactly mimic those of the listener's head nods 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 prerecorded animation of a speaker. This condition is applied whenever only one participant showed up in the experiment.

Table 5.16 shows the manipulation of the virtual speaker in each condition. In the listener's view (through the HMD), the virtual speaker's behavior follows the combination in the table.

In all the four conditions, the virtual speaker blink randomly and have slight eye movements. Its lip movements are driven by the amplitude of the speech of the real speaker.

Figure 5.5: Two Participants Are Doing the Experiment

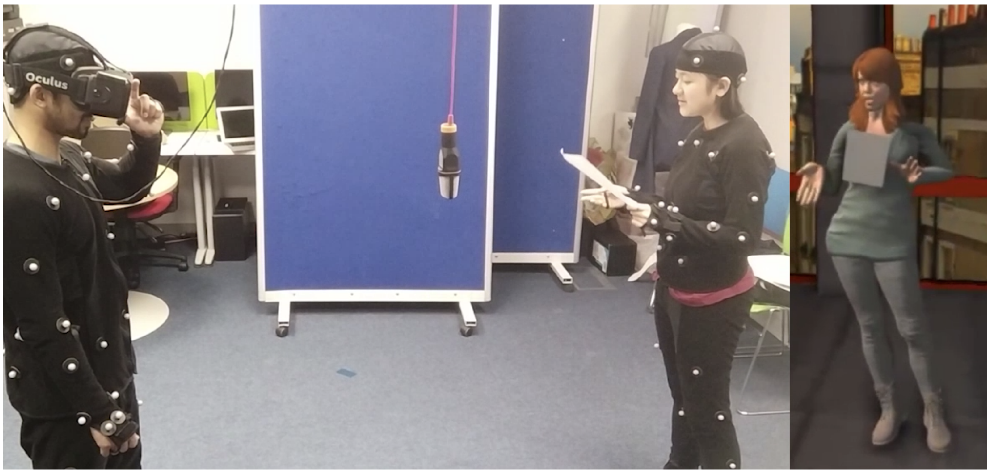


Figure 5.6: The Real Listener, Speaker and the Virtual Avatar of the Speaker

Table 5.16: The Manipulation of the Virtual Speaker by Condition

Condition	Head Movement	Speech	Body Movement
Mimic	Listener's in 4s delay	Speaker's	Speaker's
Playback	Previous listener's	Speaker's	Speaker's
Natural	Speaker's	Speaker's	Speaker's
Recording	Pre-Recorded	Pre-Recorded	Pre-Recorded

5.2.3 Pairing Participants

The experiment requires 1 or 2 participants to be present. Each participant has to take part in two experimental trials. As shown in Figure 5.7, each participant is asked to first be the listener in their first part, then the speaker in their second part. In each experiment trial, the previous participant and the current participant are presented at the same time. The previous participant is in their latter part as the speaker, and the current participant is in their first part as the listener.

Table 5.17 shows the number of participants in each condition. 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 familiar with the words they

Table 5.17: Number of Participants in Each Condition

Mimic	Playback	Natural	Recording	Total
14	9	12	19	54

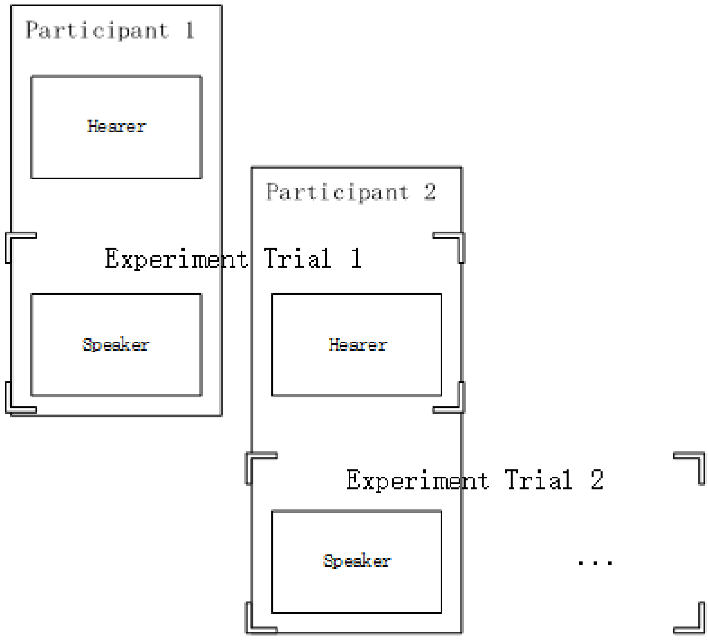


Figure 5.7: The Procedure Flow of Experiment II

would need to deliver to the listener. In the case of only one participant presents in the experiment, e.g., the very first experiment trial or one participant is not showing up, the recording condition is done since only one participant is needed in this condition.

5.2.4 Effectiveness of the Avatar Speaker

The composite measure of the agent's effectiveness from the "Digital Chameleons" study is replicated by taking the mean response to the 4 agreement questions (how much the participant agreed with the speaker's persuasive message), 12 items on impressions of the speaker, and 7 questions on the speaker's social presence. Based on the result from the first experiment, the null hypothesis is:

Null Hypothesis 2 *The speaker does not differ in their effectiveness across Mimic, Playback, Natural, and Recording conditions.*

And it is supposedthat the speaker will be the most effective in the natural condition. Thus, thealternative hypothesis is:

Hypothesis 1 *The effectiveness of the speaker is higher in the natural condition than the other three conditions (Mimic, Playback, and Recording).*

Table 5.18: Results for the Components of the Speaker's Effectiveness Measures

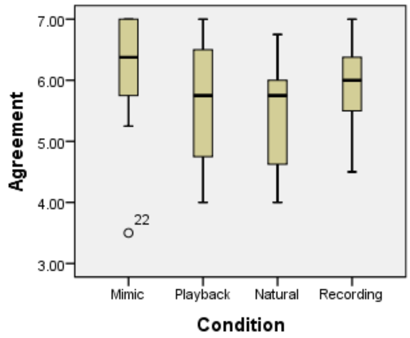
Measure	# items	α	Condition	Mean	SD	F	Sig.
Agreement	4	0.894	Mimic	6.196	0.97	1.933	0.136
			Playback	5.583	1.12		
			Natural	5.417	0.89		
			Recording	5.921	0.70		
Impression	12	0.923	Mimic	5.911	0.94	2.448	0.074
			Playback	5.574	1.13		
			Natural	5.535	0.82		
			Recording	5.047	0.86		
Social presence	7	0.660	Mimic	4.969	0.97	3.105	0.035
			Playback	5.444	0.78		
			Natural	4.643	1.01		
			Recording	4.256	1.14		
Effectiveness	23	0.909	Mimic	5.676	0.80	2.366	0.082
			Playback	5.536	0.88		
			Natural	5.238	0.72		
			Recording	4.972	0.80		

5.2.5 Results

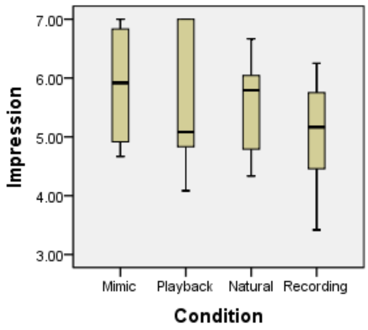
5.2.5.1 Original Composition

One-way ANOVA is used to test the effectiveness of the speaker within different conditions. Table 5.18 shows the results for the four components of the speaker's agreement, social presence, impression, and effectiveness measures; α is the Cronbach's alpha for the composite measure; F and Sig. value is from one-way ANOVA. The ratings for the agreement, impression, social presence and effectiveness of the speakers are relatively higher than those measures in the first experiment for the evaluation of IVAs. No significant effect is found with the agreement, impression and effectiveness of the speakers. However, there are significant difference in speaker's social presence across conditions (F=3.105, p=0.035).

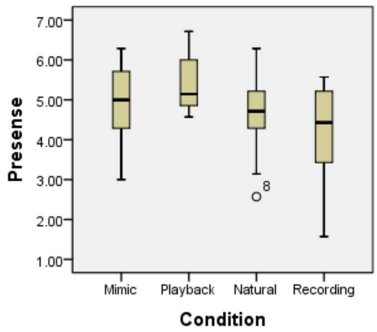
A Tukey HSD pairwise posthoc tests are applied on the four measurements. The results suggest there is no significant difference in agreement or effectiveness between the four conditions. But there are significant effects found with the impression and social presence. The impression of the speaker in the mimic condition is higher than in the recording condition (Mean Diff.=0.86, SE=0.32, p=0.05). The social presence in the playback condition is higher than in the recording condition (Mean Diff.=1.19, SE=0.41, p=0.03). Figure 5.8 is the Boxplots for the four



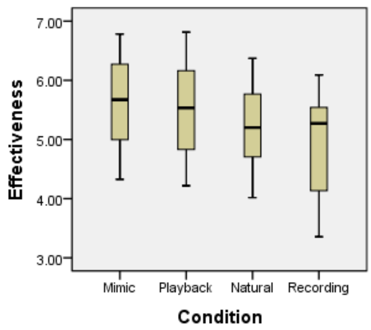
(a) Agreement



(b) Impression



(c) Social Presence



(d) Effectiveness

Figure 5.8: The Boxplots of the Components of the Speaker's Effectiveness Measures

components of the speaker's effectiveness measures.

5.2.5.2 Composition with Principal Component Analysis

A principal component analysis on the 23 items revealed six factors that together accounted for 75.4% of the total variance. All factors had eigenvalues of over 1. The item has a factor loading which its absolute value is greater than 0.3 was considered as a valid item in that factor. The factor loadings of each item and the reliability of each factor are shown in Table 5.19. The first factor can be described as effectiveness because it included 22 items of the total 23 items. The second factor can be described as ignorant because it included more negative loadings of competent, informed, credible, sincere, and positive loadings of 4 agreement items and likable. The third factor can be described as the regulation because it included valuable, workable, needed, honest, credible, and isolated as positive loadings, and friendly as negative loadings. The fourth factor can be described as virtual presence because it included the current situation, meet again, embarrassed as positive loadings, warm, modest as negative

Table 5.19: Factor Components and Loadings in Speaker-Listener Interaction.

Factor	α	Eigenvalue	Item	Factor Loading
Effectiveness (22 Items)	0.905	8.67	Agreement	0.55
			Valuable	0.42
			Workable	0.32
			Needed	0.36
			Friendly	0.73
			Likeable	0.76
			Honest	0.67
			Competent	0.53
			Warm	0.77
			Informed	0.43
			Credible	0.52
			Modest	0.71
			Approachable	0.77
			Interesting	0.65
			Trustworthy	0.78
			Sincere	0.73
			Enjoyed	0.75
			Current Situation	0.55
			Isolated	-0.31
			Meet Again	0.62
			Comfortable	0.73
			Cooperative	0.66
Ignorant (9 Items)	0.784	2.72	Agreement	0.38
			Valuable	0.46
			Workable	0.48
			Needed	0.56
			Likeable	0.30
			Competent	-0.54
			Informed	-0.69
			Credible	-0.62
			Sincere	-0.31
The Regulation (7 Items)	0.534	2.08	Valuable	0.61
			Workable	0.60
			Needed	0.40
			Friendly	-0.33
			Honest	0.37
			Credible	0.34
			Isolated	0.30
Social Presence (5 Items)	0.464	1.62	Warm	-0.37
			Modest	-0.42
			Current Situation	0.61
			Meet Again	0.49
			Embarrasseed	0.35
Embarrassed (3 Items)	-0.346	1.21	Agreement	-0.56
			Honest	0.30
			Embarrassed	0.61
Isolated (2 Items)	0.113	1.04	Isolated	0.74
			Embarrassed	-0.38

Table 5.20: The Result of One-way ANOVA on The Factors between Conditions.

Factor	Condition	Mean	SD	F	Sig.
Effectiveness	Mimic	0.47	0.96	2.501	0.070
	Playback	0.24	1.09		
	Natural	-0.06	0.86		
	Recording	-0.42	0.96		
Ignorant	Mimic	0.04	1.00	0.442	0.724
	Playback	0.28	1.21		
	Natural	-0.23	0.76		
	Recording	-0.02	1.06		

loadings. The fifth factor can be described as embarrassed because it included embarrassed and honest as positive loadings and agreement as negative loading. The last factor can be described as isolated as it included isolated as positive loading and embarrassed as negative loading. The scores for each factor were calculated for each participant using a regression method.

The reliability test suggests that only the first two factors are valid ($\alpha > 0.7$). The One-way ANOVA tests on the first two factors are conducted. Table 5.20 is the result of One-way ANOVA on the factors between conditions. No difference was found on effectiveness or ignorant between conditions. The pairwise T-tests suggest that the effectiveness

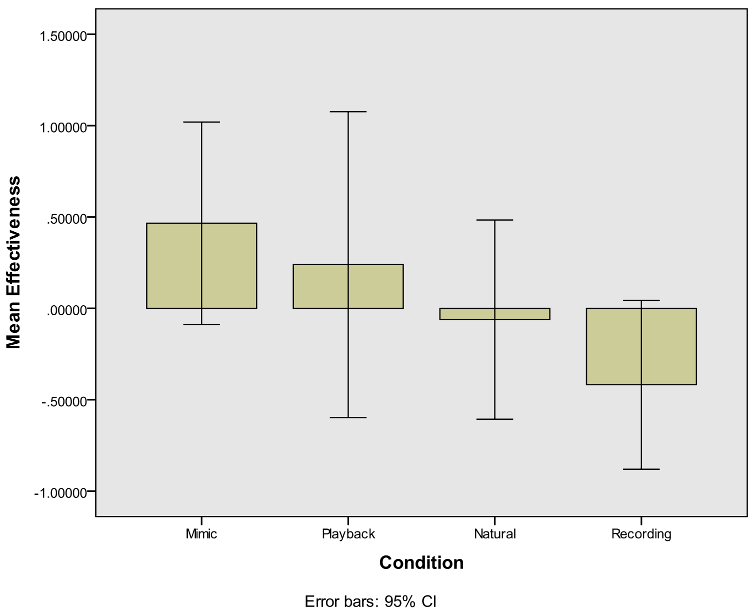


Figure 5.9: The Error-Bar Chart for the Factor of Effectiveness in the Four Conditions.

of the speaker in the mimic condition is significantly higher than in the recording condition (t_{31}=2.48, p=0.019). Figure 5.9 is the error-bar chart for the factor of effectiveness in the four conditions. However, the Tukey HSD posthoc tests suggest that the effect is weak, p=0.056.

5.2.6 Discussion

The ANOVA tests on the effectiveness of speakers across conditions show a nonsignificant effect of speaker's head nodding behavior on their effectiveness. Although the impression and social impression of speakers are found significantly higher in the mimic condition or playback condition than the recording condition, they are not the effects of mimicking behavior over natural behavior. Instead, these effects indicate that real human speakers are more effective than IVAs. Thus, the Null Hypothesis 2 which suggests no difference between speaker's effectiveness across conditions cannot be rejected. There is no significant difference in agreement or effectiveness between the mimic, playback, natural, or recording conditions. Therefore, the alternative Hypothesis 1 is rejected. The effectiveness of the speaker in the natural condition is not significantly higher than any other condition.

The PCA results suggest something different. The pairwise T-tests results show the factor of effectiveness is higher in the mimic condition than in the recording condition. However, Tukey HSD posthoc tests suggest that no significant difference in the factor of effectiveness between conditions. And no significant difference in other factors between conditions as well.

These results are quite surprising to us. In this experiment, [here is]we have a natural condition where the listener would see the exact movement of the real speaker. We expected that the effectiveness of the speaker would be different in the natural condition, either higher or lower. However, the fact is no significant difference has been detected. The PCA results in the first experiment suggest this insignificant effect should not be the result of the perceptibility of the head movement manipulation in each condition. On the other hand, this insignificant effect strongly suggests the "Digital Chameleons" effect is not robust and has some implicit limitations.

5.2.7 Conclusion

By mapping speaker-listener pairs into the immersive virtual environment, the effectiveness of the speaker in different conditions with different head behaviors is tested. The results suggest that the effectiveness of the speaker are not significantly different across conditions.

5.3 Summary

This chapter described two experiments that tried to replicate the "Digital Chameleons" effect. The first experiment used IVAs as speakers the same as the original study. The second one used participant embodied in a virtual avatar act as a real speaker. the effectiveness of the speakers in different conditions with different head behaviors is compared. However, no significant difference has been found across these conditions. Both of the experiments failed to replicate the "Digital Chameleons" effect.

Chapter 6

Evaluating the Coordination of Head Nods

In the previous chapter, the "Digital Chameleons" paradigm is evaluated with two experiments. The results suggested the speaker's effectiveness does not change significantly whether the speaker mimics the listener's head nods or not. This chapter investigates natural human behavior in these experiments.

In this chapter, the head-nodding behavior collected in these two experiments is analyzed and try to understand the natural head nods in dialogue. The head nods are analyzed from different perspectives, including the descriptive statistics of head nods, the frequency analysis of head nods, and the coordination of head nods in dialogue.

6.1 Head Nods in the IVA Speaker and Listener Interaction

In the first experiment, the effectiveness of IVAs between conditions 1)mimic – the IVAs mimicking listener's head nods 2) playback – the IVAs using the previous listener's head nods, 3) recording – and the IVAs playing a prerecorded animation is compared. The results suggest a null difference over the effectiveness of the IVAs with different head nods. And it does not result from the perceptibility of the speaker's different head-nodding behaviors across conditions. The listeners did observe the differences of the speakers' head-nodding behaviors in different conditions. How did listeners respond to these different head nods of IVAs? The listeners' head nods and their relation with IVAs' head nods are investigated using motion analysis techniques. Maximum head nodding, total head nodding and coordination of head nods between speaker and listeners for each condition are checked. Maximum head nodding is used to check participants' attention. Normal head nodding should be below 10 degrees. Participants with head nodding larger than 10 degrees means that they are looking at the floor or cell at certain timing instead looking at the speaker. This measure indicates the attractiveness of the speech. Total head nodding is used to check participants' overall head nodding movements during the speech. Larger total head nodding suggests participants' being more active. The coordination of head nodding between speakers and listeners indicates the relationship of speakers' and listener's head nodding during the speech.

6.1.1 Descriptive Statistics of Head Nods

The listeners' head nods are analyzed first with linear statistical methods. In this part, The listeners' max head nodding and total head nodding are checked. Participants' head movements are recorded at approximately 60 Hz. The max head nodding is determined by the maximum value (in degrees) of deviation from the straight-ahead position. The total head nodding is the sum of participants' total head nodding in the 2 minutes talk. Since the rating difference on the effectiveness of the IVAs across conditions is not observed, a null hypothesis is introduced:

Null Hypothesis 3 *There is no difference in participant's max or total*

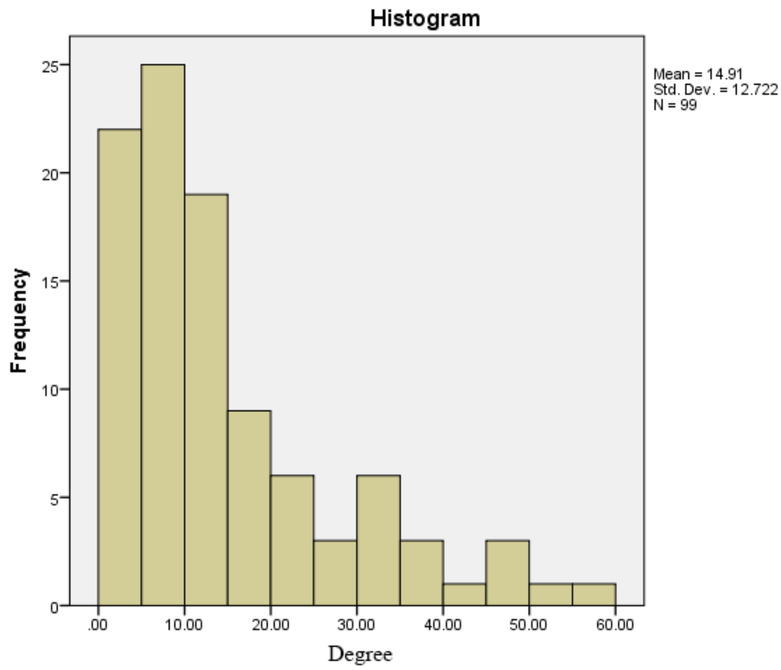


Figure 6.1: Histograph of Maximum Head Nodding

Table 6.1: The Descriptive Statistics for Participant's Maximum Head Nodding in Different Agent's Behavior Condition

Condition	Mean	Standard Deviation	95% Conf. Int.	
			Lower	Upper
Mimic	15.95	13.27	11.39	20.51
Playback	15.65	12.97	10.9	20.42
Recording	13.1	12.06	8.82	17.38
Total	14.9	12.72	12.37	17.45

head-nodding between the different conditions varies on the participant's gender, participant's order, and agent's behavior (mimic or playback or recording).

6.1.1.1 Maximum Head Nodding

Table 6.1 is the descriptive statistics for participant's maximum head nodding in different agent's behavior condition. The maximum head nodding of each participant is calculated. This measure is strongly positively skewed (Figure 6.1). A logarithmic transformation is applied to the raw data. The Shapiro-Wilk test shows that the logarithmic maximum head-nodding movement does not deviate significantly from a normal distribution (p=0.189). A GLMM analysis is used with three fixed effects (participant's gender, participant's order, and agent's behavior condi-

Table 6.2: Model Comparison of GLMM for the Logarithmic Maximum Head Nodding

Distribution & Link	Interaction	Random Intercept	BIC
Normal Identity	No	No	246.952
Normal Log	No	No	86.308
Normal Log	Yes	No	91.656
Normal Log	No	Yes	72.660

Table 6.3: Test of Fixed Effects for the Logarithmic Maximum Head Nodding

Source	F	df1	df2	Sig.
Corrected Model	2.89	4	94	0.026
Condition	0.472	2	94	0.625
Order	3.584	1	94	0.061
Gender	7.05	1	94	0.009

tion) and logarithmic maximum head-nodding movement as the target ; partcipant's order is used for repeated measure; subject is taken as the random effect. Table 6.2 shows the Bayesian Information Criterion (BIC) for model selection for the logarithmic maximum head nodding. The table suggests the model using a log normal link function, with no fixed effects interaction and included random intercept has the least BIC of 72.660. This model is selected for the analysis of the logarithmic maximum head nodding.

Table 6.3 shows the fixed effects in GLMM on the logarithmic maximum head nodding. The results suggest no model can be built to predict the logarithmic maximum head nodding by condition or order; However,

Table 6.4: Estimates of Fixed Coefficients for the Logarithmic Maximum Head Nodding

Item	Coefficient	Std. Error	t	Sig.	95% Conf. Int.	
					Lower	Upper
Intercept	0.7	0.068	10.26	<0.001	0.565	0.836
Cond.=1	0.004	0.056	0.071	0.994	-0.107	0.115
Cond.=2	0.05	0.059	0.851	0.397	-0.067	0.168
Cond.=3	0	.	.	.	.	.
Order=1	0.079	0.042	1.893	0.061	-0.004	0.162
Order=2	0	.	.	.	.	.
Gender=1	0.221	0.083	2.655	0.009	0.056	0.387
Gender=2	0	.	.	.	.	.

Table 6.5: Estimates of Covariance for the Logarithmic Maximum Head Nodding

Item	Estimate	Std. Error	Z	Sig.	95% Conf. Int.	
					Lower	Upper
Residual(Order=1)	0.244	0.089	2.733	0.006	0.119	0.5
Residual(Order=2)	0.266	0.084	3.166	0.002	0.143	0.494
Random(Intercept)	0.063	0.02	3.188	0.001	0.034	0.116

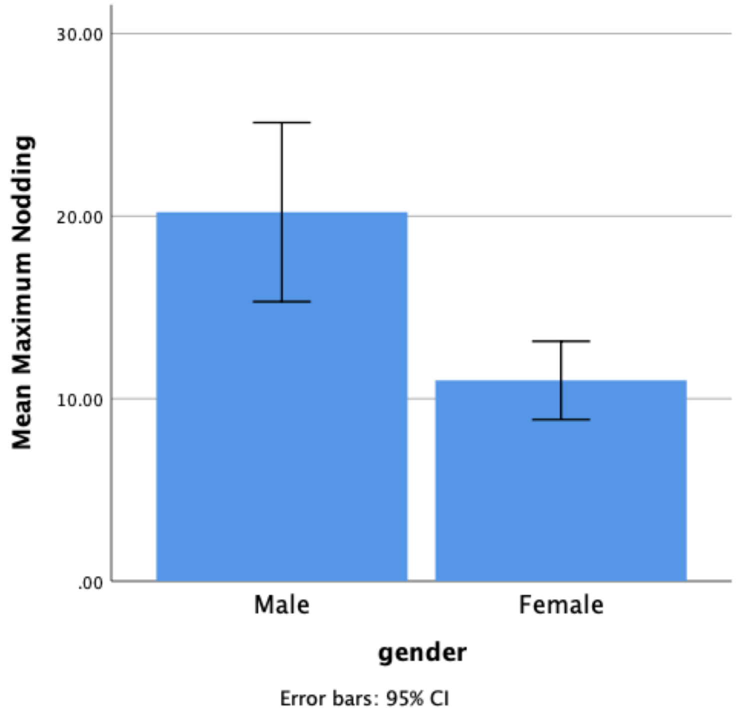


Figure 6.2: Gender Difference in Maximum Head Nodding

a GLMM can be built to predict the logarithmic maximum head nodding of the participant with participant's gender (p=0.009). There is significant fixed effects of the logarithmic maximum head nodding on participant's gender. Table 6.4 shows the estimates of the fixed coefficients for the logarithmic maximum head nodding. Table 6.5 shows the estimates of covariance parameters for the logarithmic maximum head nodding. There are variability of different order on the logarithmic maximum head nodding. There is a significant random intercept effect. This suggests that there are unmeasured variables interactions for which the logarithmic maximum head nodding of the participant appears random.

Pairwise comparison of the logarithmic maximum head nodding by participant's gender is shown in the Table 6.6. The logarithmic maximum head nodding is higher with male particpants than female participants

Table 6.6: Pairwise Comparison of the Logarithmic Maximum Head Nodding by the Participant's Gender

Gender Pair	Contrast	Std. Error	t	df	Adj. Sig.
Male - Female	0.529	0.203	2.604	94	0.011

Table 6.7: The Descriptive Statistics for Participant's Logarithmic Total Head Nodding in Different Agent's Behavior Condition

Condition	Mean	Standard Deviation	95% Conf. Int.	
			Lower	Upper
Mimic	10.48	0.747	10.23	10.74
Playback	10.46	0.76	10.18	10.74
Recording	10.3	0.83	10.0	10.59
Total	10.41	0.78	10.26	10.57

(Mean Diff.=0.529, p=0.011). Figure 6.2 illustrates the gender difference in maximum head nodding.

6.1.1.2 Total Head Nods

Since the participant's total head nodding is quite large, a logarithmic transformation is applied to the data. Table 6.7 is the descriptive statistics for participant's logarithmic total head nodding in different agent's behavior condition. Same as the maximum head nodding, the total head nods also do not fit normal distribution and are positively skewed. A Shapiro-Wilk test shows that the logarithmic total head nods did not deviate significantly from a normal distribution (p=0.191).

A GLMM analysis is run with three fixed effects (participant's gender, participant's order, and agent's behavior condition) and participant's logarithmic total head nods as the target; subject as the random factor. Table 6.8 shows the Bayesian Information Criterion (BIC) for model selection for the logarithmic total head nodding. The table suggests the model using a log normal link function, with no fixed effects interaction

Table 6.8: Model Comparison of GLMM for the Logarithmic Total Head Nodding

Distribution & Link	Interaction	Random Intercept	BIC
Normal Identity	No	No	243.785
Normal Log	No	No	-196.672
Normal Log	Yes	No	-174.778
Normal Log	No	Yes	-230.671

Table 6.9: Test of Fixed Effects for the Logarithmic Total Head Nodding

Source	F	df1	df2	Sig.
Corrected Model	1.754	4	94	0.145
Condition	0.262	2	94	0.77
Order	6.459	1	94	0.013
Gender	0.215	1	94	0.644

Table 6.10: Estimates of Fixed Coefficients for the Logarithmic Total Head Nodding

Item	Coefficient	Std. Error	t	Sig.	95% Conf. Int.	
					Lower	Upper
Intercept	2.331	0.015	156.543	<0.001	2.301	2.360
Cond.=1	-0.006	0.01	-0.616	0.54	-0.026	0.014
Cond.=2	0.0	0.01	0.016	0.978	-0.021	0.021
Cond.=3	0	.	.	.	.	.
Order=1	0.019	0.008	2.541	0.013	0.004	0.034
Order=2	0	.	.	.	.	.
Gender=1	0.009	0.019	0.463	0.644	-0.029	0.047
Gender=2	0	.	.	.	.	.

and included random intercept has the least BIC of -230.671. This model is selected for the analysis of the logarithmic total head nodding.

Table 6.3 shows the fixed effects in GLMM on the logarithmic total head nodding. The results show no model can be built to predict the logarithmic total head nodding by condition or gender; However, a GLMM can be built to predict the logarithmic total head nodding of the participant with participant's order (p=0.013). There is significant fixed effects of the logarithmic total head nodding on participant's order. Table 6.10 shows the estimates of the fixed coefficients for the logarithmic total head nodding. Table 6.11 shows the estimates of covariance parameters for the logarithmic total head nodding. There are variability of different order on the logarithmic total head nodding. There is a significant random

Table 6.11: Estimates of Covariance for the Logarithmic Total Head Nodding

Item	Estimate	Std. Error	Z	Sig.	95% Conf. Int.	
					Lower	Upper
Residual(Order=1)	0.049	0.063	0.789	0.43	0.004	0.593
Residual(Order=2)	0.252	0.079	3.204	0.001	0.137	0.466
Random(Intercept)	0.004	0.001	4.183	<0.001	0.003	0.007

Table 6.12: Pairwise Comparison of the Logarithmic Total Head Nodding by the Participant's Order

Order Pair	Contrast	Std. Error	t	df	Adj. Sig.
1 - 2	0.201	0.079	2.556	94	0.012

intercept effect. This suggests that there are unmeasured variables interactions for which the logarithmic total head nodding of the participant appears random.

Pairwise comparison of the logarithmic total head nodding by participant's order is shown in the Table 6.12. The logarithmic total head nodding is higher for participants in the first trial than the second trial (Mean Diff.=0.201, p=0.012).

The listener's total head nodding is also compared with the IVA's total head nodding. A one-sample T-test is run comparing the logarithmic agent's total head nods with the logarithmic participants' total head nods. The result suggests participants' total head nods are significantly lower than IVAs' total head nods. For male agent, $t(99) = -12.428, p < 0.001$; for female agent, $t(99) = -9.132, p < 0.001$. Figure 6.3 shows the boxplot of the logarithmic participants' total head nods. The total head nods of the male agent (actor) are represented by the dash-dot line and the female agent (actor) by the dotted line. The plot shows that the total head nods of the male agent are higher than the female agent and higher than the 95% upper confidence interval for the mean of participant's total head nods.

6.1.2 Head Nods Coordination

The head nods coordination between IVA and participant across conditions is analyzed with cross recurrence quantification analysis (xRQA). Since IVAs' head nods are copied from the participant with 4 second delay in the mimic condition and IVAs' head nods are an exact replay of the previous participant in the playback condition, the coordination of head nods between IVA and listener is expected to be the highest in the mimic condition, then playback condition, and lowest in the recording condition. Thus, the hypothesis is:

Hypothesis 2 *The coordination of head nods between IVA and listener is highest in the mimic condition, higher in the playback condition, and lowest in the recording condition.*

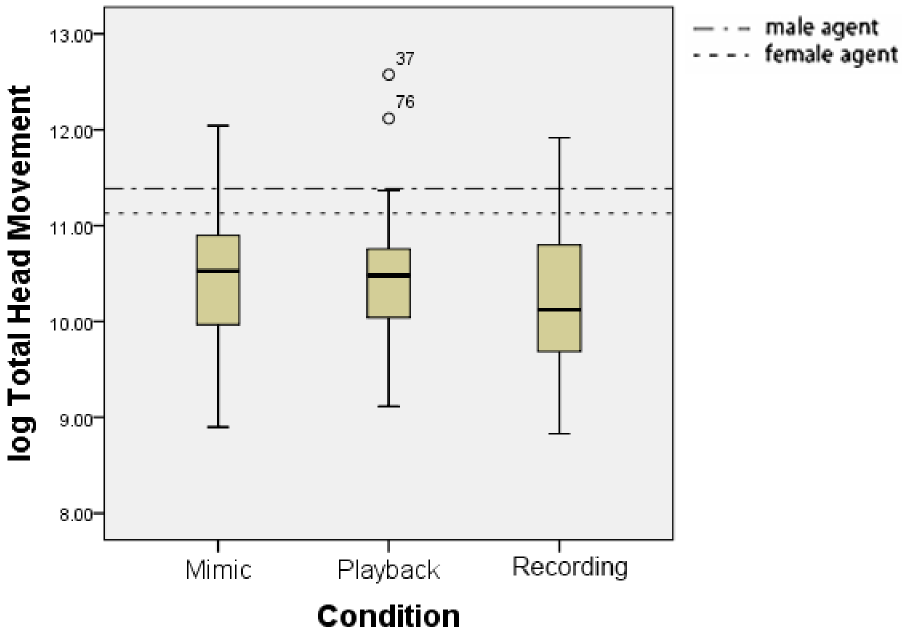


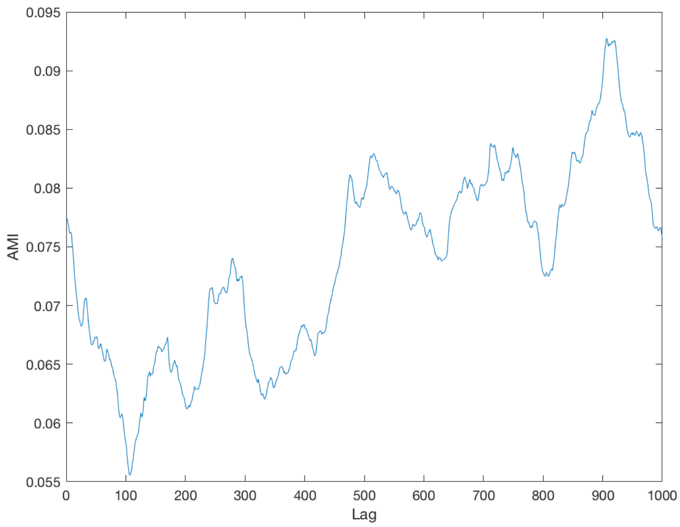
Figure 6.3: The Boxplot of Participants' Log Total Head Nods. $Mean_{mimic} = 10.483, SD_{mimic} = 0.747; Mean_{playback} = 10.46, SD_{playback} = 0.760; Mean_{recording} = 10.299, SD_{recording} = 0.832$. The 95% upper confidence interval for mean is 10.74 for mimic and playback condition, 10.59 for recording condition. The logarithmic total head nods for male agent is 11.3846, for female agent is 11.1274.

6.1.2.1 Parameters for the Cross Recurrence Quantification Analysis

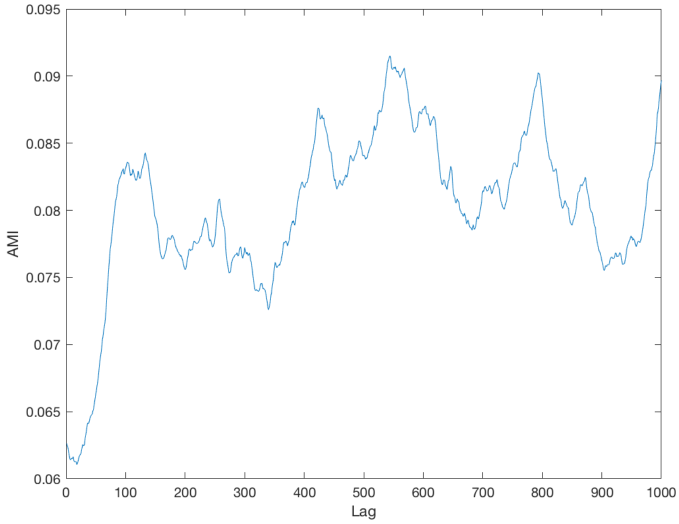
To perform the xRQA, we need to confirm the parameters for the xRQA need to be confirmed first. The steps are as follow:

First, the phase space needs to be reconstructed by selecting a fit number for the parameters – Embedding Dimension and Delay. As noted in Section 3.3.3, it is better to fix the parameters when comparing across people or groups.

The delay could be computed with the Average Mutual Information (AMI) function. For each pair of interactants, a fluctuation plot was produced. The first delay, which comes to a minimum of AMI should be chosen. Figure 6.4 is two examples of the fluctuation plot of delay. The variation between interactional pairs is huge, M=11.32, SD=18.16. We cannot chooseHowever, an optimal delay for all the interactional pairs is hard to find. As noted in Section 3.3.3, the choice of delay is not critical, and short delay is better for noise reduction, Thus, a minimal delay of 1 data frame (around 15ms) for all the interactional pairs is used.



(a)



(b)

Figure 6.4: Examples of Fluctuation Plot of Delay

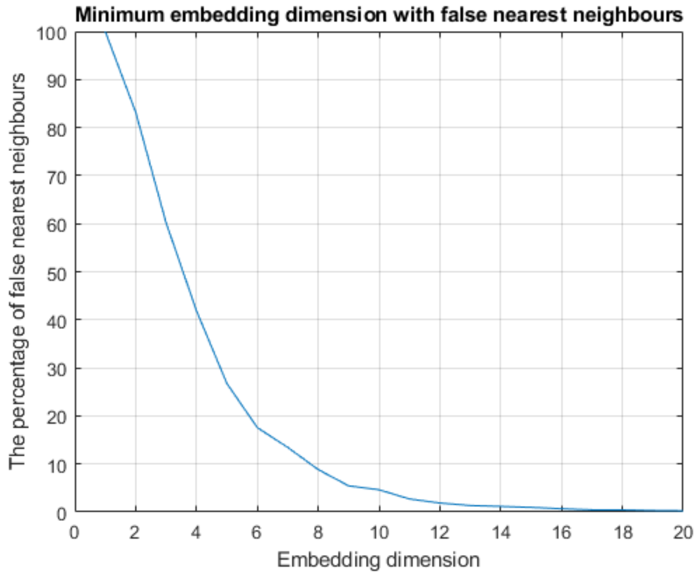


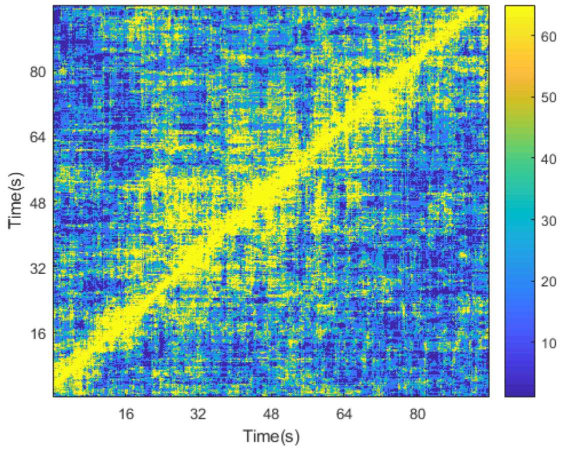
Figure 6.5: The Plot of Embedding Dimension on %FNN

The embedding dimension is calculated with the False Nearest Neighbor (FNN) function. Similarly, the first embedding dimension, which comes to a minimum of FNN should be chosen. For all the interactional pairs, similar plots are produced (Figure 6.5). The figure shows a decreasing curve of %FNN by embedding dimension. As noted in Section 3.3.3, 10-20 is a good number of embedding dimensions for the physiology system. Since an embedding dimension of 10 produced enough small %FNN, the embedding dimension of 10 is taken for all the interactional pairs.

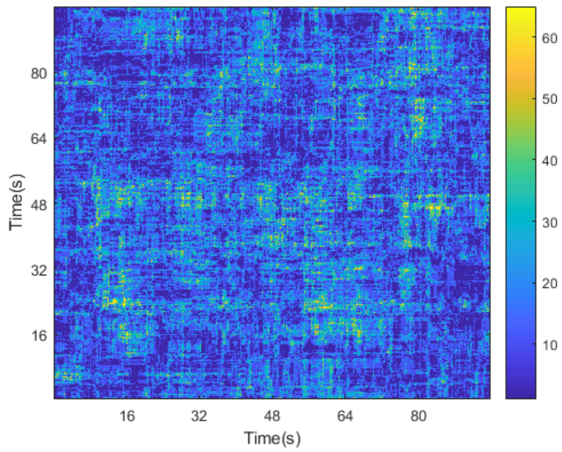
The next step is to find the right radius. A right radius should produce a %REC around 0.5% to 5%. Having tested a few different radius, a radius=50 is found suitable for all the interactional pairs.

6.1.2.2 Colored Recurrent Plots

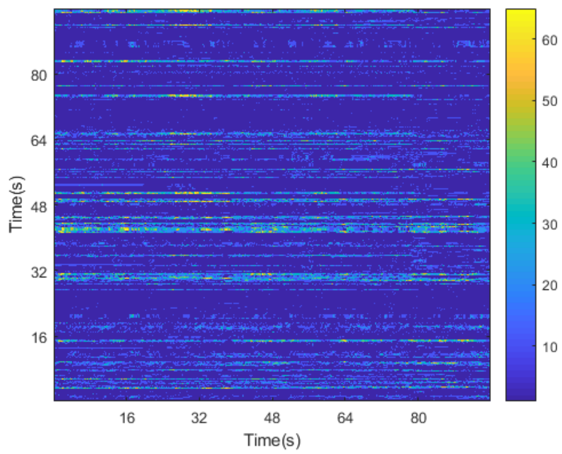
The xRQA for all the interactional pairs is run with the parameters: Embedding Dimension=10, Delay=1, Radius=50. The recurrent plots is summed to produced the colored recurrent plot by conditions. Figures 6.6 are the colored recurrent plots for coordination of head nods between IVAs and listeners in different conditions. In the figure, X axis shows the transformed participants' (listeners') head nodding intensity by interaction time. Y axis shows the transformed IVAs' (speakers') head nodding intensity by interaction time. As expected, mimicry behavior produces



(a) Mimic condition



(b) Playback condition



(c) Recording condition

Figure 6.6: The Colored Recurrent Plots in the Mimic, Playback and Recording Conditions

Table 6.13: The Descriptive Statistics for %REC, %DET and LMAX in Different Agent's Behavior Condition

item	Condition	Mean	Standard Deviation
%REC	Mimic	2.52	2.31
	Playback	1.19	0.83
	Recording	0.23	1.75
	Total	1.34	1.75
%DET	Mimic	99.04	0.82
	Playback	98.6	1.1
	Recording	96.38	1.73
	Total	98.0	1.73
LMAX	Mimic	10366.69	426.38
	Playback	111.62	41.56
	Recording	41.21	16.04
	Total	3787.95	4975.18

a lot of square recurrence patterns (Figure 6.6a). And a yellow digonal line with an offset about 4 second in the direction of Y axis can be seen in the figure, which indicates the manipulation in the mimic condition that the IVA's head nodding copies the participant's with 4 second delay. The experimental manipulation guarantees this. The recurrence pattern in the playback condition has more vertical lines (Figure 6.6b). And the recurrence pattern in the recording condition has many horizontal lines and much more fragmented (Figure 6.6c). This is because the contrast head movements between the IVAs and the participants. The IVAs' head movements are exaggerated since its animation is produced by the confederates who is pretending to give a speech to an imaginary listener. Thus the IVAs' head keeps contantly moving. On the contrary, the participants' head are relatively steady. The colored recurrent plot produced with an exaggerated movement and a steady movement would produce mostly blue areas. Occasionally, there are moments that the IVAs are doing a relatively steady movement while the participants' are still in the steady state. At these moments, they are highly synchronized and showing as pale blue horizontal lines in the colored recurrent plot.

6.1.2.3 Analysis of xRQA Outputs

Three main xRQA outputs are used for the analysis of head nodding coordination: %REC, %DET, LMAX. In terms of speaker and listener's head nodding, %REC is the rate of the speaker's nods repeated by the listener

Table 6.14: Model Comparison of GLMM for the xRQA Outputs

Distribution & Link	Interaction	Random Intercept	BIC
Normal Identity	No	No	351.439
Normal Identity	Yes	No	341.9
Normal Identity	Yes	Yes	346.281
Normal Log	No	No	358.335

which indicates the synchrony between the speaker and the listener head nods; %DET is the rate of synchrony moments over the whole speech which suggests the predictability of the listener's head nods giving the speaker's head nods; LMAX is the longest during that the speakers nods repeated by the listner which suggests the stabability of the synchrony process. Table 6.13 is the descriptive statistics of the xRQA outputs.

A GLMM analysis is used with three fixed effects (participant's gender, participant's order, and agent's behavior condition) and %REC, %DET, LMAX as the target ; partcipant's order is used for repeated measure; subject is taken as the random effect. Table 6.14 shows the Bayesian Information Criterion (BIC) for model selection for the xRQA outputs. The table suggests the model using a log normal link function, with fixed effects interaction and no random intercept has the least BIC of 341.9. This model is selected for the analysis of the xRQA outputs.

Table 6.15 shows the fixed effects in GLMM on the xRQA outputs. The results suggest there are GLMMs can be built to predict the %REC with agent's head nodding behavior condition ($p < 0.001$); %DET with agent's head nodding behavior condition ($p < 0.001$) ,participant's order (p=0.002) and participant's gender (p=0.016); LMAX with agent's head nodding behavior condition ($p < 0.001$), particpant's order (p-0.036) and participant's gender (p=0.002). There is also interaction effects found in the test of %DET (agent's head nodding behavior condition by participant's order, p=0.001) and LMAX (agent's head nodding behavior condition by participant's gender, p=0.036).

The experimental manipulation in the mimicry condition guarantees that LMAX (≈10000) is greater than the other two conditions (LMAX < 200) since it ensures perfect mimicry at a 4-second delay.

Pairwise contrasts is shown with the Table 6.16, Table 6.17 and Table 6.18. The results suggests that %REC is the highest in the mimic condition, than in the playback condition, and is the least in the recording condition; %DET is the higher in the mimic and playback condition than

Table 6.15: Test of Fixed Effects for the xRQA Outputs. The mimic condition is excluded for the test of LMAX because the LMAX in mimic condition is extremely (100x) higher than the other two conditions.

Target	Source	F	df1	df2	Sig.
%REC	Corrected Model	5.356	9	87	<0.001
	Condition	16.437	2	87	<0.001
	Order	0.018	1	87	0.893
	Gender	0.044	1	87	0.834
	Condition*Order	0.295	2	87	0.745
	Condition*Gender	1.1	2	87	0.337
	Order*Gender	1.136	1	87	0.289
%DET	Corrected Model	15.844	9	87	<0.001
	Condition	49.153	2	87	<0.001
	Order	9.93	1	87	0.002
	Gender	5.984	1	87	0.016
	Condition*Order	7.474	2	87	0.001
	Condition*Gender	0.638	2	87	0.531
	Order*Gender	0.066	1	87	0.797
LMAX	Corrected Model	24.026	6	55	<0.001
	Condition	92.344	1	55	<0.001
	Order	4.625	1	55	0.036
	Gender	10.384	1	55	0.002
	Condition*Order	0.449	1	55	0.506
	Condition*Gender	4.645	1	55	0.036
	Order*Gender	1.552	1	55	0.218

Table 6.16: Pairwise Comparison of the %REC

Condition Pair	Contrast	Std. Error	t	df	Adj. Sig.
Mimic - Playback	1.289	0.383	3.363	87	0.001
Mimic - Recording	2.149	0.381	5.634	87	< 0.001
Playback - Recording	0.860	0.41	2.10	87	0.039

Table 6.17: Pairwise Comparison of the %DET

Contrast Pair	Contrast	Std. Error	t	df	Adj. Sig.
Mimic - Playback	0.387	0.278	1.393	87	0.167
Mimic - Recording	2.614	0.276	9.477	87	< 0.001
Playback - Recording	2.228	0.297	7.507	87	< 0.001
Order(2 - 1)	0.745	0.236	3.151	87	0.002
Male - Female	0.578	0.236	2.446	87	0.016

Table 6.18: Pairwise Comparison of the LMAX

Contrast Pair	Contrast	Std. Error	t	df	Adj. Sig.
Mimic - Playback	10243.6	12.0	855.6	87	< 0.001
Mimic - Recording	10309.6	11.9	869.4	87	< 0.001
Playback - Recording	66.0	12.76	5.173	87	< 0.001
Order(1 - 2)	265.96	10.186	26.11	87	< 0.001
Male - Female	38.486	10.126	3.801	87	< 0.001

in the recording condition, however, no significant difference between the mimic condition and the playback condition; %DET is also higher for male than for female; %DET is higher in the second trial than in the first trial; for LMAX, it is the highest in the mimic condition, than in the playback condition, and the least in the recording condition; it is significantly higher for male participants than for female participants; it is significantly higher for the first trial participants than the second trial participants

6.1.3 Discussion

The result of participant's max and total head-nodding suggested that the null hypothesis 3 can be partly rejected. That is, although agent's behavior or participant's order does not affect participant's head movements, female participants moves their heads significantly less than male participants (Figure 6.2). This gender difference in head movements is also reported in the "Digital Chameleons" study. It suggests that female participants are more focused when listening to the speech. On the other hand, the primary comparison between participant's and agent's total head nods shows that participants as listeners make only limited head movements. This is strong evidence of the first problem addressed in Section 5.1.5. That is, a speaker usually moves their heads more frequently and in a wider variety of ways than listeners. The pairwise contrast of participant's total head nodding between the first trial and the second trial shows that participants' head moved less in the second trial than in the first trial(Table 6.12). This might because participants are more focused in the second trial. The speaker will departure from normal behavior if he/she simply copies a listener's head movements. It breaks the balance of initiative typical of ordinary conversation.

Perhaps more interesting are the results of the cross recurrence quantification analysis of head movements. The results suggest that the hypothesis 2 cannot be rejected. They show some apparent differences in behavior patterns for the interactions in three different conditions. This shows the head movements of the agents and participants had highest levels of repetition (%REC, %DET and LMAX) in the mimicry condition (as expected); medium %REC, higher %DET and lower LMAX in playback condition; lowest %REC, %DET and LMAX in recording condition.

Since in the mimicry condition, the agent repeats the participant's head movements with a 4s delay, it has the highest %REC, %DET, and LMAX. More importantly, the playback condition has higher %REC, %DET, and LMAX than in the recording condition. In the playback condition, the xRQA correlates two separate participants' head movements, and as listener head movements are generally reduced, this can lead to a higher repetition of (non) movement. In contrast to this in the recording condition, the subjects of xRQA are the head movements of a participant as a listener and the head movements of a confederate as a speaker. As noted, speakers move their heads significantly more than listeners, and this means the chance level of matching non-moving heads is much lower. The implication of this is that natural interaction is actually characterized by *low* levels of speaker-listener mimicry because of asymmetries in both the level of head movements and their functions. This is consistent with the literature on human interaction but incompatible with an automatic mimicry model.

6.1.4 Conclusion

The results from head nods analysis provide no evidence of differences between conditions in participant's overall responsiveness, as indicated by gross measures of their concurrent head movements. Analysis of motion capture data suggests the presence of gender differences when participants interact with the agents. The human speakers (recording) move much more than their mimicker or non-mimicker (playback) counterparts. Furthermore, cross recurrence quantification analysis on the captured motion data reveals differences in coordination of interactions in the mimic, playback, and recording conditions. People repeat the human speaker's movements systematically *less* than the mimicker (trivially) or non-mimicker (playback) (non-trivially).

This experiment replicates the original "Digital Chameleons" experiment in a new, more realistic behavioral setting, provides a critical 'natural' baseline/control condition not present in the original study and provides a new, more detailed analysis of the degree of nonverbal coordination observed in this setting. The results demonstrate a problem with this paradigm, namely that it misconstrues the underlying reciprocal dynamics of natural interaction i.e., speakers do not behave at all like listeners.

There is a limitation of this experiment. Although the third condition – recording is used to test the persuasiveness of natural interaction, it is only a prerecorded animation. A more conclusive test of the importance of mimicry requires manipulation of live communications.

6.2 Head Nods in Virtual Avatar Speaker-Listener Interaction

As noted in the first experiment, a replication of the "Digital Chameleons" paradigm with live interactions is needed. Thus the second experiment is conducted. In this experiment, the effectiveness of the virtual speaker is compared between conditions 1) mimic – the virtual avatar of the speaker mimicking listener's head nods 2) playback – the virtual avatar of the speaker using the previous listener's head nods 3) natural – the virtual avatar of the speaker use the speaker's own head nods 4) recording – and the virtual avatar of the speaker playing a prerecorded animation. The results suggest a null difference over the effectiveness of the virtual avatar of the speaker with different head nods.

In this experiment, the data of the virtual speaker and listener's head nods, as well as the data of the real speaker and listener's head nods is collected. Finally, the head nods of the virtual interactant pair and the real interactant pair is analyzed respectively.

6.2.1 Descriptive Statistics of Head Nods

With the recorded head movement data, the differences of standard deviation and total head nods between conditions are tested. Based on the results from the first experiment, the standard deviation and total head nods of listeners is supposed to be not vary across conditions. Thus, the

Table 6.19: The ANOVA tests for Listeners' SD and Total Head Nodding in Different Agent's Behavior Condition

item	Condition	Mean	SD	F	df1	df2	Sig.
SD Head Nodding	Mimic	2.34	1.45	0.275	3	50	0.843
	Playback	2.17	1.40				
	Natural	2.34	2.30				
	Recording	2.82	2.49				
	Total	2.48	2.02				
Total Head Nodding	Mimic	8805.6	6236	0.451	3	50	0.718
	Playback	7738.6	6877				
	Natural	8292.3	7807				
	Recording	10914	9606				
	Total	9255.9	7893.6				

null hypothesis is:

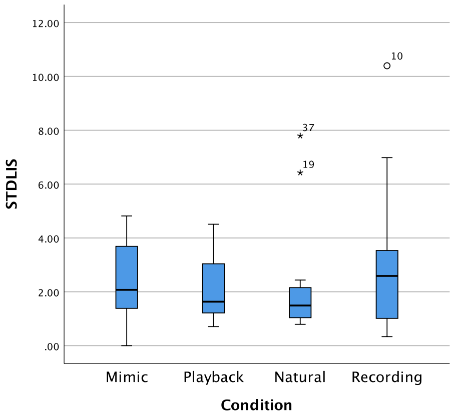
Null Hypothesis 4 *There is no difference in participants' standard deviation or total head-nodding between the different conditions varies from the speakers' behavior (mimic, playback, natural, recording).*

6.2.1.1 Results

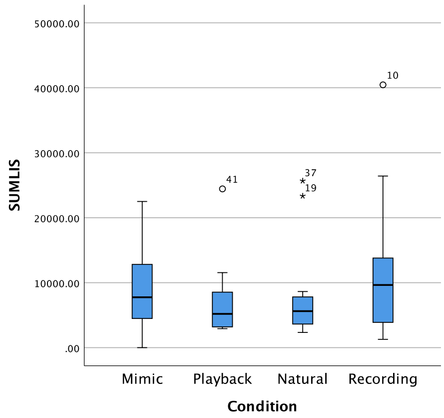
The listeners' head nods as well as real and virtual speakers' head nods are analyzed ANOVA. No significant effect is found with listeners' SD or total head nods (Figure 6.7a 6.7b). Real and Virtual Speakers' SD and total head nods are higher in recording condition than the other conditions (Figure 6.7c 6.7d 6.7e 6.7f). The bars in green or brown in the figures means they are identical within the condition. No other significant effect was found. Furthermore, the paired t-test shows no difference between speaker and listener's SD or total head nods in mimic, playback, or natural condition. However, in the recording condition, the speakers' SD or total head nods are significantly higher than the listeners (Table 6.20). No correlation is found between speakers and listeners' SD or total head nods.

6.2.2 High and Low Frequency Head Nods

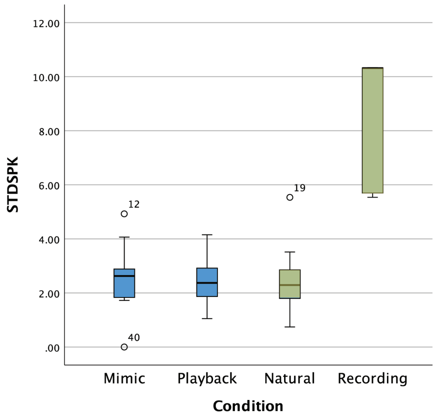
Recent research by Hale et al. (2019) suggested that people communicate with high and low frequency head nods, listeners produce more high-frequency head nods than speakers. On the contrary, Healey et al. (2013) suggested that speakers nod more than primary addressees. The results



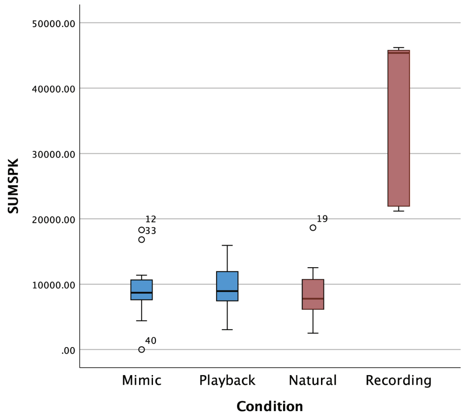
(a) SD Head Nods for Listeners (The range of nodding movements of listeners)



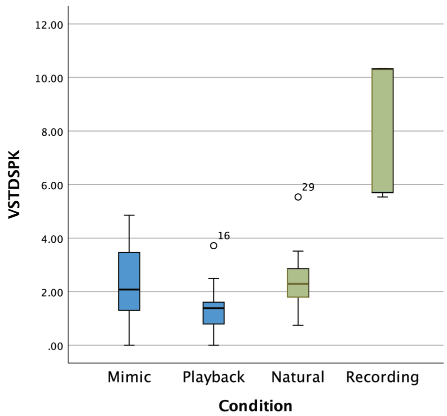
(b) Total Head Nods for Listeners (The overall nodding movement of listeners)



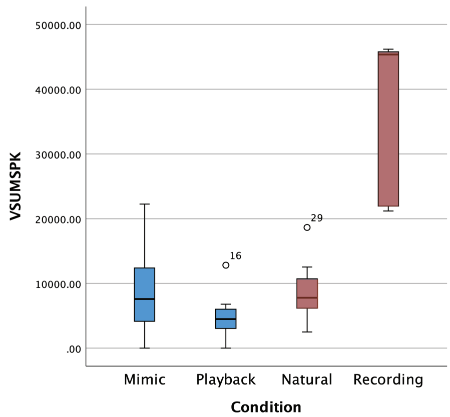
(c) SD Head Nods for Real Speakers (The range of nodding movements of speakers)



(d) Total Head Nods for Real Speakers (The overall nodding movement of speakers)



(e) SD Head Nods for Virtual Speakers (The range of nodding movements of virtual speaker characters)



(f) Total Head Nods for Virtual Speakers (The overall nodding movement of virtual speaker characters)

Figure 6.7: The SD and Total Head Nods for Listener and Speaker

Table 6.20: Paired t-tests: The Difference between Speaker and Listener's Head Nods in Different Condition

Item	Condition	Mean Diff.	SD	Sig.
SD Head Nods	Mimic	0.22	1.35	0.539
	Playback	0.21	1.54	0.692
	Natural	0.14	2.25	0.826
	Recording	6.0	3.3	< 0.001
Total Head Nods	Mimic	371	5772	0.814
	Playback	1688	5560	0.389
	Natural	447	7790	0.846
	Recording	26979	14704	< 0.001

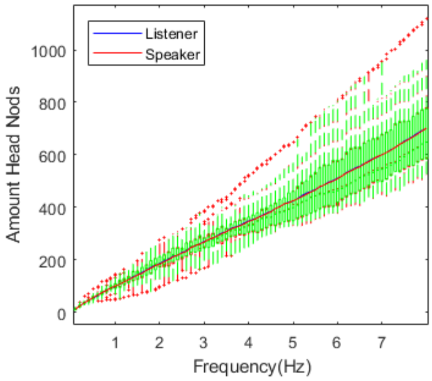
on total head nods in our first experiment also suggested speakers nod more than listeners. Do speakers nod more than listener in all frequency range? To investigate the question, the hypothesis is proposed::

Hypothesis 3 *Speakers nod more than listeners in all frequency range across conditions.*

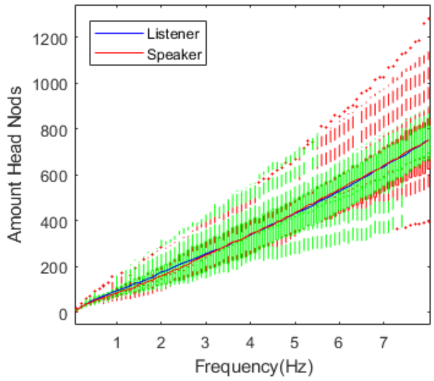
6.2.2.1 results

The number of head nods is counted for every pair of participants. Figure 6.8 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 specific low pass filter. The blue lines in the figure is the mean amount of head nods for listeners (the green boxes) in each frequncy band. The red line is the mean amount of head nods for speakers (the red boxes) in each frequncy band. The boxes are taken in the resolution of 0.1 Hz. The figures suggest that listeners nod more than speakers in high frequency band.

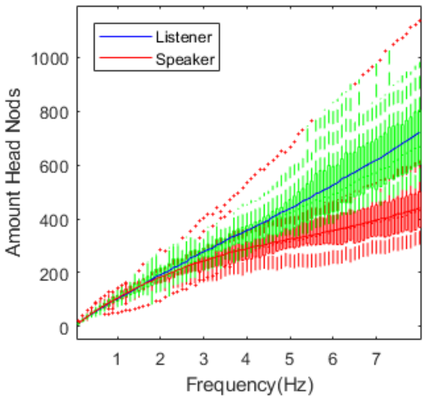
The mean difference of the number of head nods is compared between the listener and speaker below the specific frequency with the paired t-test. The result suggests that for the virtual pair of speaker and listener, there is no significant difference of the number of head nods under the condition of mimic and playback. However, the amount of real listeners' head nods are siginificantly higher than the real speakers' 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



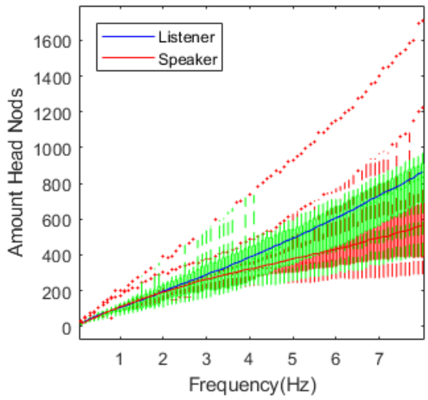
(a) The Amount of Head Nods for the Virtual Speakers and Listeners in the Mimic Condition



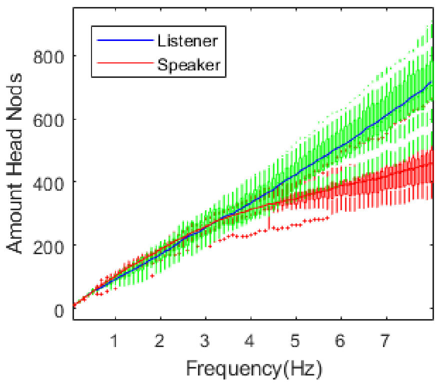
(b) The Amount of Head Nods for the Virtual Speakers and Listeners in the Playback Condition



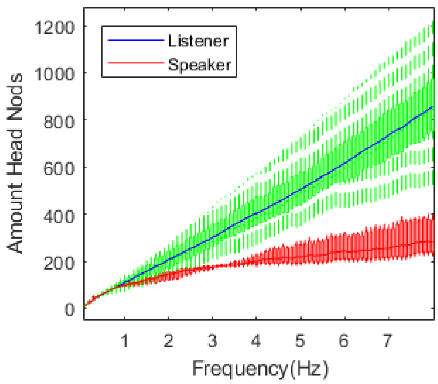
(c) The Amount of Head Nods for the Real Speakers and Listeners in the Mimic Condition



(d) The Amount of Head Nods for the Real Speakers and Listeners in the Playback Condition



(e) The Amount of Head Nods for the Speakers and Listeners in the Natural Condition



(f) The Amount of Head Nods for the Speakers and Listeners in the Recording Condition

Figure 6.8: Boxplots of the Cumulative Amount of Head Nods for the Virtual and Real Speaker-Listener Pair. It indicates the distribution of the amount of head nods for speakers and listeners in different frequency band from 0-8Hz, resolution in 0.1 Hz.

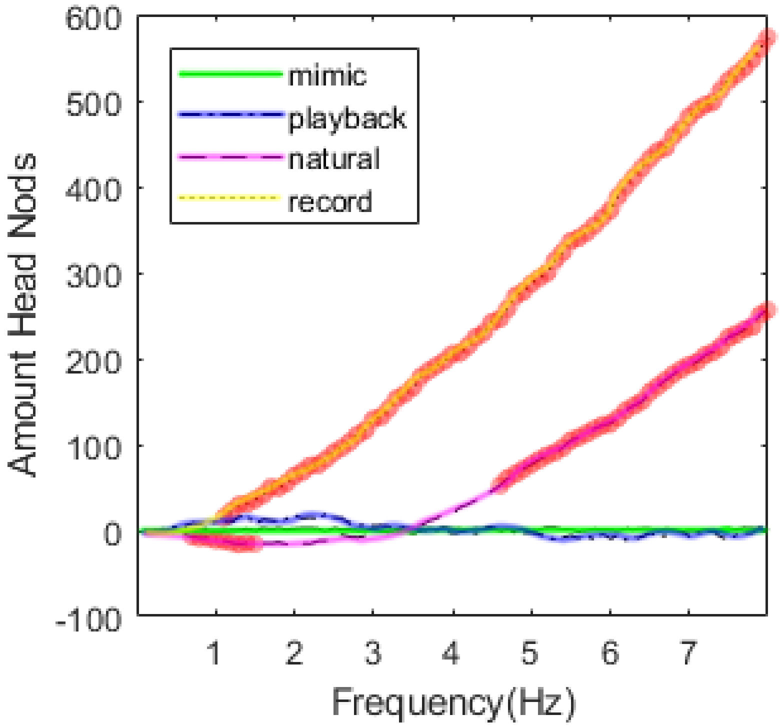


Figure 6.9: Cumulative Mean Difference of the Amount of Head Nods for the Virtual Listener-Speaker Pair. A t-test is used to compare the amount of head nods between the virtual listener and speaker in every 0.1 Hz frequency band from 0 to 8 Hz. The red dots in the graph indicate the points are a significant level of which $p < 0.05$.

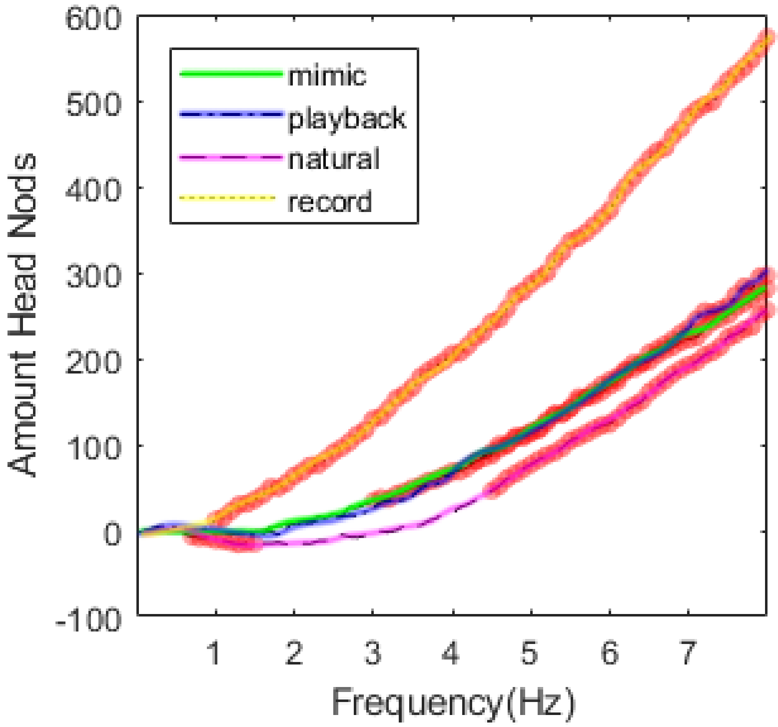


Figure 6.10: Cumulative Mean Difference of the Amount of Head Nods for the Real Listener-Speaker Pair. A t-test is used to compare the amount of head nods between the real listener and speaker in every 0.1 Hz frequency band from 0 to 8 Hz. The red dots in the graph indicate the points are a significant level of which $p < 0.05$.

the recording condition, the listener nodded significantly more than the speaker beyond 1 Hz. Figure 6.9 and 6.10 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 at a significant level of which $p < 0.05$.

6.2.3 Head Nods Coordination

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

Hypothesis 4 *Coordination of the speaker-listener's head nods is higher than chance in all conditions.*

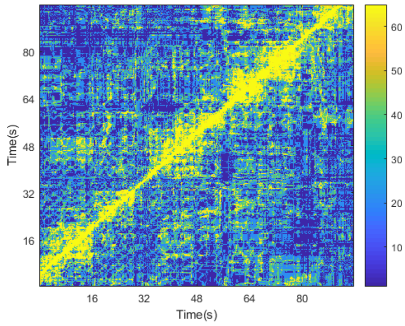
6.2.3.1 Colored Recurrent Plots

xRQA is run for all the virtual and real interactional pairs with fixed parameters: Embedding Dimension=6, Time Lag=1, Radius=50, Non-normalized. The fixed parameters ensured that the parameters are kept as the controlled variables; the value of the parameters was picked to ensure no floor or ceiling effect for the xRQA outputs; not normalize the data to reduce the effect of non-movement. Figure 6.12 and Figure 6.11 are the colored recurrence plot (CRP) for the head nods of the virtual and real speaker-listener pair in different conditions.

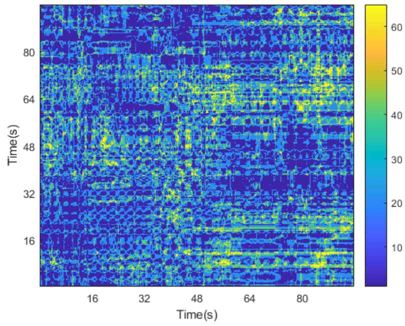
In different conditions, different patterns is shown. The CRPs of the virtual and real speaker-listener pair for mimic and playback condition are quite diverse, while in the natural and recording conditions, they are similar to each other. This is because of the experimental manipulation – the virtual and real speaker's head nods are the same in the natural and recording condition. The virtual pairs with the mimic (Figure 6.11a) and playback (Figure 6.11b) condition are more coordinated (more yellow dots in the CRP) than the other conditions. They show different coordinating patterns, e.g., there is a long diagonal line in the CRPs of the mimic condition, which is not seen in the CRPs of the playback condition. The diagonal line has a small offset in the Y-axis, which indicated the 4s delay mimicry manipulation of the virtual speaker's head-nodding. The CRP of the recording (Figure 6.12d) condition shows the least coordination (least dots) of the speaker-listener pair. However, the difference cannot be easily identified between the CRPs of the mimic (Figure 6.12a), playback (Figure 6.12b) and natural (Figure 6.12c) conditions with the real pairs.

6.2.3.2 Analysis of xRQA Outputs

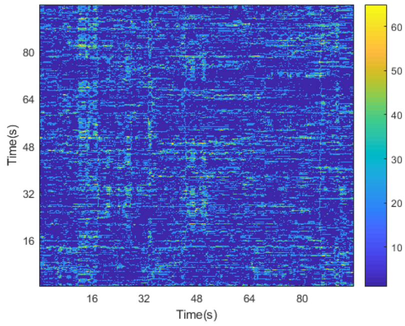
The quantification outputs of the xRQA consist of the %REC, LMAX, and %DET for all the virtual and real speaker-listener pairs. Figure 6.13 is the boxplots for those xRQA outputs by condition. The horizontal red lines are the chance level of these measures with the 95% confidence interval. The %REC, LMAX, and %DET for virtual and real speaker-listener pairs between conditions with ANOVA. The result suggests there



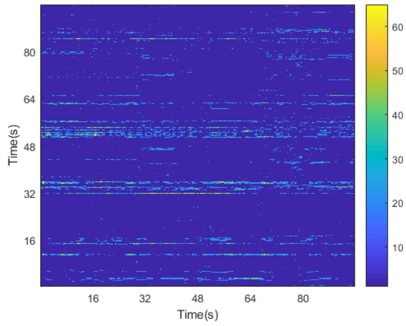
(a) Mimic condition



(b) Playback condition



(c) Natural condition

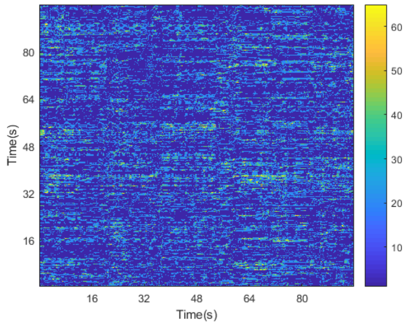


(d) Recording condition

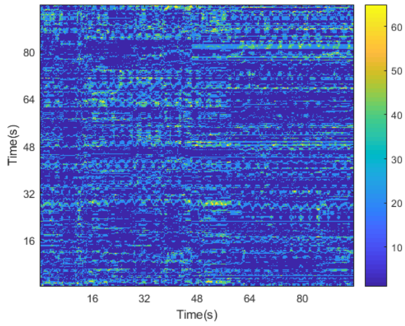
Figure 6.11: The Colored Recurrent Plots for Virtual Speaker-Listener Pair. The virtual pairs with the mimic (Figure 6.11a) and playback (Figure 6.11b) condition are more coordinated (more yellow dots in the CRP) than the other conditions. They show different coordinating patterns, e.g., there is a long diagonal line in the CRPs of the mimic condition, which is not seen in the CRPs of the playback condition. The diagonal line has a small offset in the Y-axis, which indicated the 4s delay mimicry manipulation of the virtual speaker's head-nodding. The CRP of the recording (Figure 6.11d) condition shows the least coordination (least dots) of the speaker-listener pair.

is a significant (p<0.001) difference between conditions on these items for the virtual and real speaker-listener pairs.

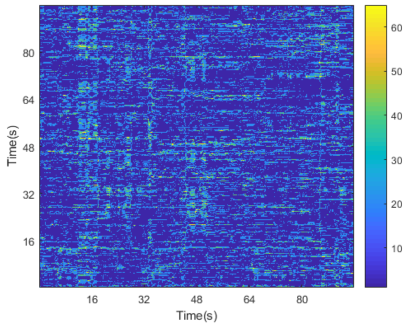
There is a main effect of speaker's behavior ($p < 0.001$) on %REC, LMAX and %DET (Table 6.21 6.22). The pairwise comparison suggests for virtual speaker-listener pairs, %REC and %DET are significantly higher in the mimic and playback condition than the natural condition, $p < 0.001$. %REC is significantly higher in the natural condition than the recording condition, $p < 0.05$. No significant difference of %REC and %DET is found between the mimic and playback condition. LMAX is significantly higher in the mimic condition than all the other three conditions, $p < 0.001$. No significant difference in LMAX is found between the



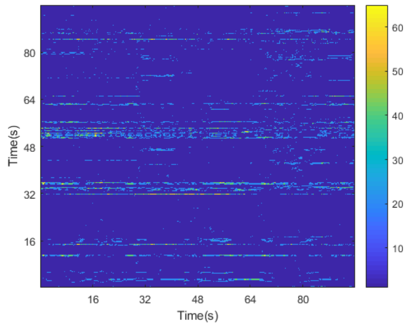
(a) Mimic condition



(b) Playback condition



(c) Natural condition

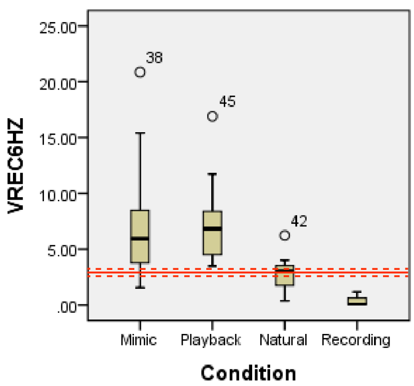


(d) Recording condition

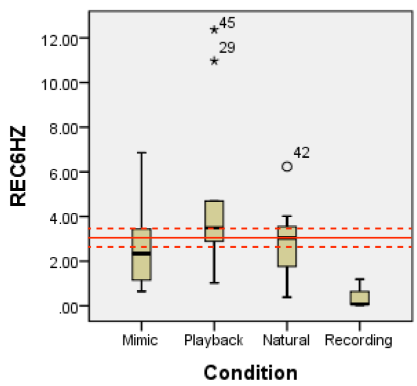
Figure 6.12: The Colored Recurrent Plots for Real Speaker-Listener Pair. The CRP of the recording (Figure 6.12d) condition shows the least coordination (least dots) of the speaker-listener pair. However, the difference cannot be easily identified between the CRPs of the mimic (Figure 6.12a), playback (Figure 6.12b) and natural (Figure 6.12c) conditions with the real pairs.

Table 6.21: ANOVA Tests for xRQA Outputs of Virtual Speaker-Listener Pair

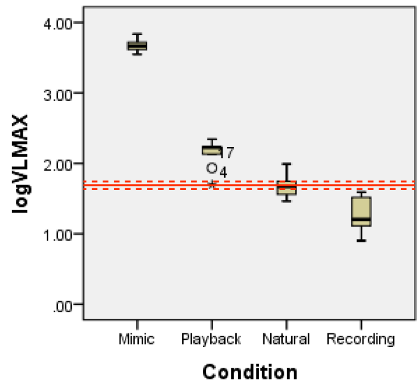
item	Condition	Mean	SE	F	df1	df2	Sig.
%REC	Mimic	7.22	0.88	16.297	3	50	<0.001
	Playback	7.52	1.1				
	Natural	2.78	0.95				
	Recording	0.32	0.76				
LMAX	Mimic	4738.57	117.1	394.504	3	50	<0.001
	Playback	150.11	146.0				
	Natural	50.75	126.48				
	Recording	20.74	100.52				
%DET	Mimic	99.31	0.70	26.226	3	50	<0.001
	Playback	99.56	0.88				
	Natural	96.72	0.76				
	Recording	92.19	0.60				



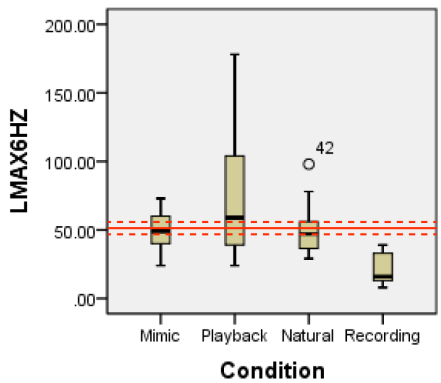
(a) %REC for Virtual Speaker-Listener Pair



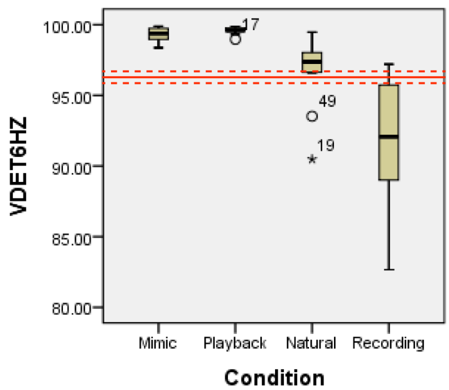
(b) %REC for Real Speaker-Listener Pair



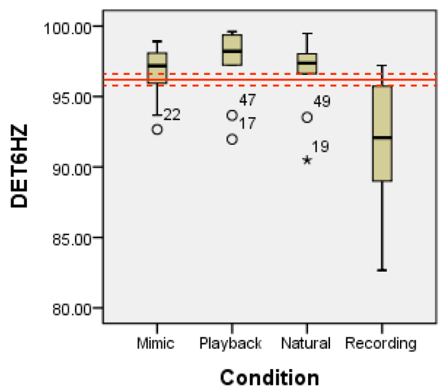
(c) logarithmic LMAX for Virtual Speaker-Listener Pair



(d) LMAX for Real Speaker-Listener Pair



(e) %DET for Virtual Speaker-Listener Pair



(f) %DET for Real Speaker-Listener Pair

Figure 6.13: Boxplots of xRQA Outputs for the Virtual and Real Speaker-Listener Pair. The horizontal red lines are the chance level of these measures with the 95% confidence interval.

Table 6.22: Cross Recurrence Quantification Analysis of Real Speaker-Listener Pair

item	Condition	Mean	SE	F	df1	df2	Sig.
%REC	Mimic	2.49	0.53	11.521	3	50	<0.001
	Playback	4.87	0.663				
	Natural	2.784	0.575				
	Recording	0.32	0.457				
LMAX	Mimic	49.14	6.57	11.521	3	50	<0.001
	Playback	78.22	8.20				
	Natural	50.75	7.10				
	Recording	20.74	5.64				
%DET	Mimic	99.74	0.80	10.14	3	50	<0.001
	Playback	97.24	0.995				
	Natural	96.72	0.86				
	Recording	92.19	0.69				

other three conditions. For the real speaker-listener pairs, %REC and LMAX are significantly higher in the playback condition than other conditions, $p < 0.01$. %REC and LMAX are significantly higher in the mimic and natural conditions than in the recording condition, $p < 0.01$. No significant effect is found between mimic and natural conditions. %DET is significantly higher in mimic, playback, and natural conditions than in the recording condition. No significant difference of %DET between mimic, playback, and natural condition is found.

Games-Howell posthoc pairwise test suggests that: for the virtual speaker-listener pairs, %REC is not significantly different from the chance level in the mimic, playback and natural condition, while it is significantly below the chance level in the recording condition, Mean Difference (MD)=2.72, p<0.001; LMAX is greater in the mimic condition than in the playback condition (MD=4588, p<0.001), and it is greater in the playback condition than in the natural condition, (MD=99.4, p<0.005). It is at about chance level in the natural condition and greater than in the recording condition (MD=32.5, p<0.001); %DET is above the chance level in the mimic condition (MD=2.75, p<0.001) and playback condition (MD=3.0, p<0.001) , 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 pairs, %REC is below the chance level in the recording condition (MD=2.72, p<0.001) while not significantly different from the chance level in the mimic, playback and natural condition; LMAX is not reliably different from chance in the mimic and natural conditions, whereas it is above the chance level in the playback condition and below the chance level in the recording condition; %DET is not significantly different from the chance level in the mimic, playback, and natural condition, while it is significantly below the chance level in the recording condition (MD=4.37, $p<0.001$).

6.2.4 Discussion

The standard deviation and total head nods is analyzed respectively for the listener and speaker. The results suggests that there is no significant difference in the standard deviation and total head nods for listeners across mimic, playback, natural, and recording conditions. Thus, the null hypothesis 4 cannot be rejected. There is also no difference in the standard deviation and total head nods for speakers across live interactions, that is, mimic, playback, natural conditions. On the other hand, the results suggest the IVA speakers in the recording condition move their head significantly more than the speakers in the other conditions. This suggests that using an actor to perform communication without the presence of a real listener is extremely wrong and not at all as the speaker would do in real conversation. The paired t-test also suggests that no significant difference in the standard deviation and total head nods between speaker and listener. Note that the total head nods are calculated based on the Equation 3.1.

A much more salient and surprising finding is the distribution of head-nodding behavior by the speaker and listener during the dialogue. In terms of the number of head nods, the results show that listeners nod 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 is observed in the other conditions. This suggests that the hypothesis 3 is partly rejected. Speakers do not nod more than listeners in all frequency range across conditions. In natural communication, the speaker and listener nod differently in the high and low frequency domain (Hale et al., 2019). Moreover, Figure 6.8f indicates that the speaker in the recording condition nod much less in the high-frequency area than the speaker in the other conditions. This is despite the fact that people performing the monologue in the recorded condition move 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 behaviors, they might be a critical 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 nonverbal performance that is very different from the natural behavior of a speaker in a live interaction.

The speaker listener's head-nodding coordination is tested by applying the one-way ANOVA to the xRQA outputs. The most obvious point about the results illustrated in Figure 6.13 is that coordination with the recorded speaker is consistently well below the 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 from 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. It 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 the head movements of listeners is paired 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 the rejection of the hypothesis 4 as well. The coordination of the speaker-listener's head nods is not higher than change in all conditions. 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 the 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 (mimic condition) or from another listener (playback condition). In contrast, 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., 2019).

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) suggested that temporally shifting the timing of hand gestures in the video away from its audio component creates an abnormal 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.

6.2.5 Conclusion

The speaker and listener's head nods in a live interaction are analyzed with four different conditions. The results suggest that speaker and listener head nods are various. In the natural interaction condition, people do not coordinate their nodding behavior more than would be expected by chance. The analysis of head-nodding behavior suggests that this is because speakers nod more in the low-frequency domain and less in the high-frequency field 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. The study also finds that fast nods are critical in the speaker-listener's coordination.

6.3 Summary

This chapter described the evaluation of the coordination of head nods in the two experiments. The speaker and listener head nods are analyzed with traditional linear statistical method, frequency analysis, and non-linear time series analysis method (xRQA). The results suggest speakers and listeners nod differently. The level of coordination of head nods between speaker and listener in the natural interaction is not higher than chance.

Chapter 7

Discussion

In this chapter, all the findings from the two experiments are discussed, such as how these experiments are connected, and how these findings are related to the literature. The discussion is divided into three sections.

The limits of the "Digital Chameleons" paradigm is first discussed. The section 7.2 tries to explain why the "Digital Chameleons" did not work in the two experiments. The section 7.3 investigates the coordination of head nods in dialogue, which is the ultimate question of the thesis. Finally, the methodological approach of this study is discussed, including the use of IVE, the replication of the "Digital Chameleons" paradigm, the evaluation of the effectiveness of the virtual speakers, and a selection of behavior analysis methods.

7.1 Summary of Hypotheses and Outcomes

At the beginning of this chapter, the list of all the hypotheses and their outcomes are presented in Table 7.1 (the hypotheses about the relationship between mimicry and effectiveness in the thesis and their outcomes), Table 7.2 (the hypotheses about head nodding in the dialogue in the thesis and their outcomes) and Table 7.3 (the hypotheses about the coordination of nods in the dialogue in the thesis and their outcomes).

Table 7.1: The Hypotheses About the Relationship Between Mimicry and Effectiveness in the Thesis and Their Outcomes

Hypothesis	Null Hypothesis 1
Question	Does head nodding mimicry improve effectiveness of an IVA?
Description	There is no difference in the effectiveness of the agent between the different conditions (mimic or playback or recording).
Outcome	Accepted
Hypothesis	Null Hypothesis 2
Question	Does head nodding mimicry improve effectiveness of the speaker with a virtual avatar?
Description	The speaker does not differ in their effectiveness across Mimic, Playback, Natural, and Recording conditions.
Outcome	Accepted

7.2 Limitations of the "Digital Chameleons" Paradigm

The "Digital Chameleons" paradigm, which has been mentioned many times in this thesis, is not successfully replicated again in the two experiments described in this thesis (Chapter 5). This adds another two failed replications along with the previous studies (Hale & Hamilton, 2016; Riek et al., 2010).

There are several possible explanations for this. For example, the survey design is bad, it may not suit for the one-way communication with IVA(Riek et al., 2010); virtual reality can not perfectly replicate the real human behavior; virtual mimicry is not the naturalist mimicry (Hale & Hamilton, 2016); or the effects of virtual mimicry varies with different tasks (Hale & Hamilton, 2016; Verberne, Ham, & Midden, 2015;

Table 7.2: The Hypotheses About Head Nodding in the Dialogue in the Thesis and Their Outcomes

Hypothesis	Null Hypothesis 3
Question	Does participant nod differently in their max or total head nodding across conditions when interacting with IVA?
Description	There is no difference in participant's max or total head-nodding between the different conditions varies on the participant's gender, participant's order, and agent's behavior (mimic or playback or recording).
Outcome	Partly rejected The maximum head nodding of the male participants is greater than the female participants (Contrast = $10^{0.529} \approx 3.38$ degrees, Std. Error = $10^{0.203} \approx 1.60$, p = 0.011).
Hypothesis	Null Hypothesis 4
Question	Does participant nod differently in their standard deviation or total head nodding across conditions when interacting with the speaker embodied with a virtual avatar?
Description	There is no difference in participants' standard deviation or total head-nodding between the different conditions varies from the speakers' behavior (mimic, playback, natural, recording).
Outcome	Accepted
Hypothesis	Hypothesis 3
Question	Does speaker nod more than listener in all frequency range across condition?
Description	Speakers nod more than listeners in all frequency range across conditions.
Outcome	Rejected Listeners nod more than speakers in the frequency between 3-8 Hz in the natural condition ($p < 0.05$).

Verberne et al., 2013). This study suggests another explanation – the behavior of the "Digital Chameleons" is deviated from the normal human behavior. Thus the "Digital Chameleons" would not be more effective than ordinary people.

7.2.1 False Condition Design

The conditions of the original "Digital Chameleons" study are pretty constrained:

Table 7.3: The Hypotheses About the Coordination of Head Nods in the Dialogue in the Thesis and Their Outcomes

Hypothesis	Hypothesis 2
Question	Does participant and IVA coordinate differently in head nodding across conditions?
Description	The coordination of head nods between IVA and listener is highest in the mimic condition, higher in the playback condition, and lowest in the recording condition.
Outcome	Accepted
Hypothesis	Hypothesis 4
Question	Does participant and the speaker embodied with a virtual avatar coordinate higher than chance in head nodding across conditions?
Description	Coordination of the speaker-listener's head nods is higher than chance in all conditions.
Outcome	Partly rejected %REC for the speaker-listener's head nods is not significantly differ from chance in the mimic, playback and natural conditions and lower than chance in the recording condition. LMAX and %DET are higher than chance in the mimic and playback conditions ($p < 0.001$), not significantly differ from chance in the natural condition, and lower than chance in the recording condition ($p < 0.001$).

The mimic condition suggests that the speaker constantly mimics the listener's head movement. This is not the case in real communication. People might only mimic other's behavior occasionally. Some other unpredictable behavior could be brought into the mimic condition from which the IVA would copy exactly the head movements in 4s delay. Participants would not constantly look at the IVA without looking around in the virtual environment while this would cause a very unnatural behavior with the mimicker IVA – giving a speech with its head looking around. This movement would not create a better social influence.

Chartrand et al. (1999; 2013) suggested that human unconsciously mimic their interaction partner and gain social influence. However, it does not mean that IVA would also gain social influence by implementing simple automatic mimicry rules. People learnt to mimic others in the social interaction (Catmur & Heyes, 2013; Heyes, 2001, 2011). There might be an automatic judgment in the human mind on when to mimic. For example, mimicry behavior can be facilitated with social intention

(Chartrand & Lakin, 2013). Furthermore, as suggested by Giorgolo and Verstraten (2008), the automatic mimic behavior might violate the fact that human nonverbal behaviors are tightly coupled with speech content and social context. IVA could possibly gain social influence by mimicry if applied with a sophisticated, intelligent algorithm. This algorithm should have the capability of social awareness to perform the mimicry task in the social interaction context.

On the other hand, in the playback condition, the IVA used the previous listener's head movement, which is also pretty weird. In the report of the original "Digital Chameleons" study (Bailenson & Yee, 2005), the playback condition was set to ensure that, in the current experiment trial, the head movements of the IVA are the same as its head movements in the previous experiment trial (mimic condition). At the same time, the IVA does not mimic the participant's head movements in the current experiment trial. This design promised the experimental control – the same head movements mimic or not mimic the participant's head movements.

However, the same as the problem noted above, the head movements of the IVA in the playback condition are disconnected from the IVA's speech content and social context. Nevertheless, comparing the effectiveness of the mimic IVA with the playback IVA is not the same as comparing mimicry behavior with natural behavior. Instead, it is the comparison between the effectiveness of one weird behavior with another odd behavior.

In the first experiment, the third condition where the IVA used prerecorded animations is added. IVA with prerecorded animation is a widely used technique in the film or game industry. It is more reasonable for the IVA speaker to use the prerecorded animations than the previous listener's movements. A significant difference in the effectiveness of the IVAs is not found in the recording condition. However, the results suggest the IVAs in the recording condition does show more 'authority' than the IVAs in the mimic and playback conditions (Figure 5.4). Furthermore, no difference is found in 'authority' between the IVAs in the mimic condition and the IVAs in the playback condition. This suggests using the IVAs with prerecorded animation is better than using the IVAs with the previous listeners' movements. It also indicates the limitation of the "Digital Chameleons" paradigm – the effect is too subtle and fragile.

7.2.2 Agents versus Avatars

In the second experiment, the natural condition where the speaker's exact movements would be tracked and mapped to the movements of the virtual avatar is added. It is believed that the virtual avatar using natural speaker's movements would have more social influence than the virtual avatar mimicking participant's head nods or using the previous participant's head nods, or the IVAs using prerecorded animations. As suggested with previous research, participants gesture more to avatars than IVAs; participants perform better when talking to avatars than talking to IVAs with prerecorded animation (Dodds, Mohler, & Bülthoff, 2011); avatar excluded people have higher levels of sadness and are less helpful and less confident than avatar included people (Kothgassner et al., 2017); avatars are more influential than IVAs in social interactions (Blascovich et al., 2002; Fox, Fox, & Hall, n.d.); avatars facilitate more attitude change (Guadagno et al., 2007). These differences are fundamentally brought in by human thoughts. Since the avatar is driven by a real person, the behaviors of avatar are no different from a real person.

However, in this study, the results suggest the virtual avatar mimicking the participant's head nods get slightly more agreement than the virtual avatar with natural movements. They further suggest avatars are slightly more effective than the IVAs using the prerecorded animations. But none of these effects are significant. This might be because the head nods behavior is highly context-dependent and serves as communicative functions. The nonconscious or automatic mimicry behaviors only consist of a small part of social interaction. Or the sample size is too small to reveal the effect. This emphasizes the subtle and fragile effect of the "Digital Chameleons" paradigm.

7.2.3 Communicative Functions of Head Nods

The weirdest part of the "Digital Chameleons" paradigm is the manipulation of the speaker's head nods. Human head nods have many communicative functions. They are not just a signal of 'Yes'. Speaker's head nods are fundamentally different from the listener's head nods. For example, speakers nod to signal the intention to continue speaking, to seek or check agreement, to express emphasis, to control and organize the interaction, to 'beat' the rhythm of the speech. On the other hand, listeners nod for ‘backchannel’, to signal understanding, agreement, or support. (Hadar et al., 1985). The number of head nods is very different between speakers and listeners (Hale et al., 2019; Healey et al., 2013). This would be elaborated in the next section.

Overall, a speaker using the listener’s head nods is fundamentally wrong in the sense of natural interaction. This manipulation would break the connection of head-nodding behavior with its context and diminish the communicative function of head nods. The nonconscious or automatic head nods mimicry may happen occasionally, but it would never exist continually without breaking the communicative functions of head nods. The “Digital Chameleons” studies failed because of the false manipulation of the speaker’s head nods.

However, it would be impossible for the experimenter to determine the main factor of the effect of mimicry if it introduced other behaviors instead of using constant imitation. This creates a dilemma for the “Digital Chameleons” study, as well as many other behavior studies. How can we manipulate behavior for experimental purpose without deviating from the natural human behavior?

7.2.4 Bidirectional Mimicry

Nonconscious or automatic mimicry is a bidirectional behavior (Chartrand & Bargh, 1999; Chartrand & Lakin, 2013). If this claim is true, it would be expected that the number of head nods from both speaker and listener to increase for the nonconscious mimicry and the feedback loop between speaker and listener. Of course, the increase of head nods is not a linear function; instead, most probably, it should be a logarithmic function. In this case, it is expected that the amount of head nods of the listener goes up along with the increasing of the number of speaker’s head nods. However, the data does not support this prediction. As suggested with Figure 6.3 and Figure 6.7, the speakers in the recording condition nods the most. On the other hand, the number of head nods of the listeners in the recording condition does not significantly differ from those in the other conditions.

Again, the explanation for this would be nonconscious or automatic mimicry only consists part of the total head nods. On the contrary, most of the head nods should serve as the communicative functions. Thus, any significant difference cannot be detected in the listeners’ head nods across conditions.

7.2.5 Conclusion

The nonconscious or automatic mimicry behaviors are infrequent in natural human behaviors. The "Digital Chameleons" always mimic the listener's head nods failed to address the communicative functions of head nods and ignored the differences of head nods between speakers and listeners. Thus, any significant effect with the "Digital Chameleons" paradigm cannot be found. That is, the virtual speakers mimicking the listeners' head nods are not more effective than the natural speakers or the IVAs using a prerecorded animations.

7.3 Coordination of Nods in Dialogue

The coordination of nods in dialogue is the prime topic of the thesis. The head nodding time-series data is collected in the experiments when evaluating the speaker's effectiveness with different head nods. The head nods data can be categorized into different groups. Overall, The head nodding data of speaker and listener pairs can be classified into 6 different situations. They are the head nods of speakers and listeners when 1) the IVA mimicking the listener's head nods 2) the IVA using the previous listener's head nods 3) the IVA using a prerecorded animation; and 4) the virtual speaker mimicking the listener's head nods 5) the virtual speaker using the previous listener's head nods 6) the virtual speaker moving the same as the real speaker. This head nods data is analyzed with linear statistical methods and nonlinear methods. The following sections explain what head nods are.

7.3.1 Collaborated Head Nods

In the theories of dialogue, dialogue is modeled from different perspectives. For example, dialogue is an interactive alignment process – it suggests dialogue partners automatically mimic each other's communicative behaviors (Lakin & Chartrand, 2003) through the perception-behavior link (Chartrand & Bargh, 1999; Dijksterhuis & Bargh, 2001); dialogue is a collaborative process – it suggests listeners make a significant contribution to speakers' utterances as co-creators or co-narrators (Bavelas et al., 2000; Krauss, 1987; Wilkes-Gibbs, 1986).

The data on the head nods of speakers and listeners broadly favors the collaborative dialogue model. That is, the head nods of the speaker and listener collaborate in the dialogue. This is demonstrated by the comparison of total head nods of speakers between virtual avatar speaker-listener interaction (dialogue) and IVA speaker-listener interaction (monologue). Speakers perform significantly more head nods in the recording condition (IVA speaker used a prerecorded animation of a confederate doing a monologue without the presence of a listener) than in the other three conditions (speakers delivered the message to the present listener) (Figure 6.7). Here it should be clarified that the contributions of the listener in the collaborative dialogue do not necessarily increase the amount of a specific behavior. On the contrary, it could also decrease the amount of a particular behavior to establish a natural interaction with the speaker collaboratively. For example, in the study, the number of head nods of the speaker significantly decreased with the presence of the listener compared to that in the absence of a listener. It could be imagined that the confederate exaggerated his/her nonverbal behaviors to effectively convey the message to the imaginary listener in his/her mind.

On the other hand, no matter of the existing of the listener (mimic, playback, natural conditions) or not (recording condition), the amount of listener's total head nods do not significantly change with the head-nodding behavior of the speaker (mimicry or not) (Figure 6.3 6.7). This goes against the prediction from the interactively aligned dialogue model.

7.3.2 Unnoticeably Fast Head Nods

Head nods can be classified by their frequency. For example, 1) slow head nods between 0.2-1.8 Hz 2) ordinary head nods between 1.8-3.7 Hz and 3) rapid or fast head nods above 3.7 Hz (Hadar et al., 1985). Listeners mainly use ordinary head nods to signal 'YES', fast head nods for synchrony, and slow/ordinary nods for other tasks.

Figure 7.1 shows the mean size (degree) for slow and fast nods. While most of the slow nods are smaller than 1.5 degrees (+/- 2 SD), most of fast nods are smaller than 0.426 degrees (+/- 2 SD). The size of fast head nod is so tiny that it is very likely not able to be noticed consciously by people. These tiny head nods may come from different participant's activities. Most of these tiny head nods are served for the synchrony

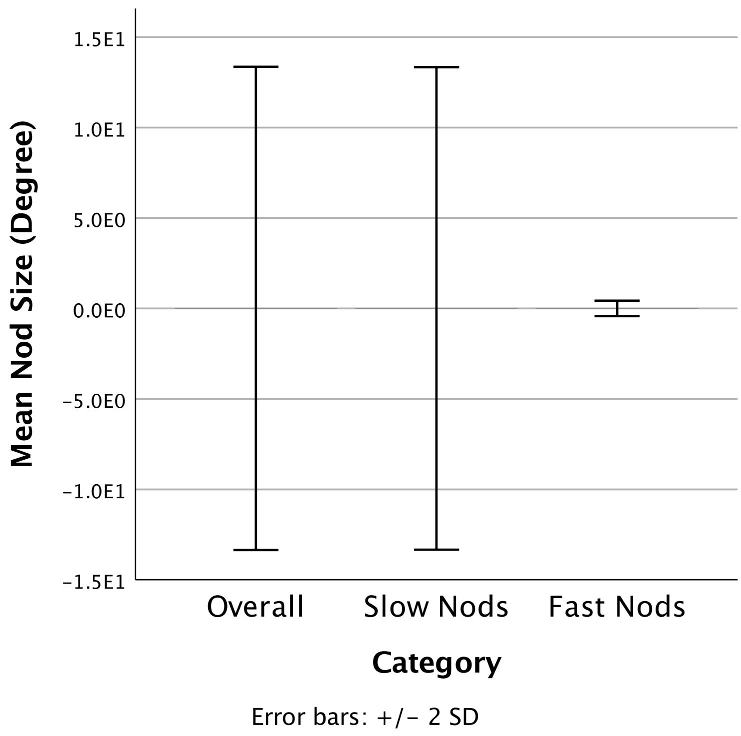


Figure 7.1: The Mean Size (Degree) for Slow and Fast Nods

function as it was suggested with Hadar (1985). On the other hand, these tiny head nods may come from the head movements leakage from the other dimensions. For example, when participants are laughing, talking, shaking, or looking around, the participants' head move largely in the other directions, such as head yaws and rolls. These movements could leak to the head pitch direction and presented as tiny head nods. In most of the cases, these head nods are ignored or categorized as other head movement behaviors. Overall, these leaked head nods only happen occasionally, and they are, in general, served for synchrony.

7.3.3 Asymmetries in Head Nods between Speakers and Listeners

In the literature, Healey et al. (2013) suggested speakers nod more than primary addressees and that this relationship varies depending on how fluent the speaker's performance is. On the contrary, Hale et al. (2019) suggested that listeners nod more in the high-frequency domain. These two studies seem to conflict with each other. However, in the study of Healey et al. (2013), the authors eliminated the high-frequency nods in their analysis. It is supposed that speakers and listeners nod differently in high and low frequency domain. If this is the case, the two studies mentioned above do not conflict with each other anymore.

Our data provided evidence supporting this hypothesis. In Section 6.2.2, the difference in the amount of head nods is compared with each frequency window range from 0 to 8 Hz in the resolution of 0.1 Hz between speakers and listeners across conditions (Figure 6.8). The results suggest in the natural interaction, speakers nod more than listeners in the low-frequency domain (0.7-1.5 Hz). On the other hand, speakers nod less than listeners in the high-frequency domain (above 3 Hz) (Figure 6.9). The results are compatible with both the results from Healey et al. (2013) and the results from Hale et al. (2019).

Nevertheless, these asymmetries in head nods between speakers and listeners have been shown in the other three unnatural conditions (mimic, playback, recording) (Figure 6.10) . It could be another evidence for the collaborated head nods hypothesis – the listener's head nods deviate from the natural behavior if the speaker's head nods are unnatural. It does not matter whether the listener is presented or not.

7.3.4 Coordinated Head Nods

The collaboration of head nods and the difference of head nods between speakers and listeners are discussed in the previous sections. In this section, the coordination of head nods between speakers and listeners in the dialogue is investigated.

As noted in Section 3.3.3, the xRQA method is used for the measure of coordination of head nods. This method measures how one movement is repeated by another movement over time. Specifically, the xRQA method has three main outputs – %REC, LMAX, and %DET. In this study, it measures how the speaker's head nods are repeated by the listener's head nods over time. Here, the %REC is the rate of the speaker's head nods repeated by the listener's; the LMAX is the length of the longest repeated head nods sequence; the %DET is the rate of deterministic (predictability) of the listener's head nods with the listener's head nods.

In this study, the %REC is found to be higher in the mimic and playback conditions than in natural condition, and lower in the recording condition. The LMAX is highest in the mimic condition, higher in the playback condition than the natural condition, and the least in the recording condition. The %DET is higher in the mimic and playback conditions than the natural condition, and the least in the recording condition. These results are consistent along with the two experiments except that in the first experiment %REC is higher in the mimic condition than in the playback condition; whereas in the second experiment, no difference is found with %REC between the two conditions. These differences can be clearly seen from the Figure 6.6 and Figure 6.11. There are more yellow dots in the CRP of the mimic condition than the CRP of the playback condition, then the CRP of the natural condition, and mostly blue dots in the CRP of the recording condition.

These results are consistent with the prediction of the hypothesis 2 which suggests the coordination of head nods between IVA and listener is highest in the mimic condition, higher in the playback condition, and lowest in the recording condition. It is trivial that the head nods of a listener would coordinate with his/her own head nods the most (the mimic condition); and would coordinate with another listener's head nods less (the playback condition); and would coordinate with a non-collaborative speaker's head nods the least (the recording condition). However, it is non-trivial that a listener repeats another listener's head nods (the playback condition) more than a listener repeats a collaborative speaker's head nods (the natural condition). The results have given a clear answer.

At the same time, the coordination of head nods in the mimic and playback conditions are above chance coordination. The coordination of head nods in the recording condition is below chance coordination. These results make sense with the collaborative dialogue view. Because without the presence of listener (recording condition), the speaker's head nods are not coordinated thus leading to lower than chance coordination.

To our surprise, the coordination of head nods in the natural interaction does not differ from the chance coordination (Figure 6.13). This result suggests that the coordination of head nods in the dialogue between speaker and listener should be no more than it would be expected by chance. This is consistent with the results reported in a previous study (Plant, 2018).

The head nods are randomly (chance) coordinated for the real speaker-listener pairs in the mimic and playback conditions as well. Figure 6.12 shows that the CRPs in these two conditions does not seem to differ much from the CRP in the natural condition. This suggests that the coordination of head nods are at the same level in the conditions where the speaker have a listener (mimic, playback, natural conditions). And it is higher than the head nods coordination in the condition where the speaker does not have a listener (recording condition) (Figure 6.13b 6.13d).

Nevertheless, the speakers and listeners's head nods in the three live interaction conditions (mimic, playback, natural conditions) are coordinated in different ways. The frequency analysis (Figure 6.10) shows that speakers nod more in the low-frequency domain than listeners. On the other hand, speakers' nods are no significant different from listeners' nods in the mimic and playback conditions. A possible explanation would be speakers are constantly monitoring listeners' nonverbal cues and adjust their head nods according to listeners' head nods. At the same time, listeners are constantly monitoring speakers' nonverbal cues and adjust their head nods according to speakers' head nods as well. In this way, speakers and listeners maintain a specific level of coordination spontaneously. And their head nods become coordinated.

7.3.5 Conclusion

Head nods are collaborative behaviors in dialogue. With the presence of the listener, the speaker nods less in the dialogue than the speaker in a monologue. Head nods can be classified into slow, ordinary, and fast head nods. Fast head nods are mostly smaller than 0.5 degrees and might be unnoticeable, but they are contributing to synchrony and should not be ignored. Speakers and listeners' head nods are asymmetric in high and low frequency domain. Specifically, speakers nod more in the low-frequency domain and less in the high-frequency domain than listeners. The coordination of head nods in the natural dialogue is no more than the chance level of head nods coordination. Speakers' and listeners' head nods become coordinated by collaboratively adjusting their head nods spontaneously.

7.4 Methodological Approach

The following sections discuss the methodological approach with an emphasis on how the selected methods contribute to the primary goals and findings of this thesis. It also reflects on the shortcomings of the methods that arise from the practices of the studies and future works that could be done to improve the evaluation of the coordination of head nods in dialogue.

7.4.1 Using Immersive Virtual Environment

The idea of using IVE in the behavior research field is not novel. It has been proved to be a promising method with a series of research (see Section 2.3.5). It is used to enable the multimodal real-time full-body interaction and has excellent experimental controllability. On the other hand, it puts a high technical requirement on the experimenter. It needs to be carefully used, especially when it is connected with a full-body motion capture system – another professional system.

Although there is nothing novel, a few reflections is brought forward after the heavy use of the IVE and the full-body motion capture system.

7.4.1.1 Design of IVE

In Section 4.3, three different ways to design the IVE are discussed: 1) a blank IVE 2) the same as the real-world laboratory 3) base on the scenario. The third design – an office environment is used for this study. Although the experiments are not run with the other two IVEs, the results suggest the appropriateness of the use of the third IVE design. However, it also can be improved if the tracking space in the lab can be matched with the IVE so that the participants could freely walk around in the IVE.

7.4.1.2 Design of Virtual Characters

As it was suggested in the literature, visual realism and behavioral realism focused on the different aspect of using IVEs (Guadagno et al., 2007; Toczek, 2016; Zibrek et al., 2019). For example, in the research looking for the psychological effects of the environment, human appearance, human race, prime, visual realism would be critical. In other research looking for the psychological effects of body ownership, place illusion, plausibility illusion, and nonverbal behavior in IVEs, behavioral realism is essential. Since there are mixed effects of the uncanny valley (Mori et al., 2012), the trade-off between the agency and tracking accuracy (Argelaguet, 2016), and the limitations of technology, the visual realism, and behavioral realism should be carefully weighed. However, using abstract virtual characters is not quite suitable for our virtual environment as well as not ideal for social interaction. The three ways in designing virtual characters are discussed in Section 4.3.1, and finally realistic virtual characters is used in the study. At the same time, a fine-grained motion capture technology (Vicon) is used to improve behavioral realism. The results suggest that the virtual characters perform well in terms of head-nodding behavior. However, improvement needs to be done with hand movements.

7.4.1.3 Full-body Motion Tracking

In order to enable the most natural interaction in the IVE, Vicon Motion Capture system is used for the full-body motion tracking. This system requires users to put on the black suits with markers on specific positions, which varies with participants' figures. It is impossible to maintain the tracking accuracy with the predefined settings of the suits. Each participant has to put on the black suits. They are asked to make a T-pose until all the markers are placed on their proper position. After that, the participant has to do a set of calibrating movements to create a tracking profile for this specific participant. This process is very time-consuming; it occupies almost 70% of the experiment time. A better full-body tracking solution needs to be implemented in future studies.

Nevertheless, this system is not able to track hand movements, which is reported to be the most unnatural part of the study. In the future, Leap Motion or data gloves can be integrated to the system to enable hand tracking for social interaction. There is also face tracking with HMD, which could be used in the future as well.

7.4.1.4 Experimental Controllability

IVE provided exceptional experimental controllability. However, this controllability has been misused in many studies. For example, in this study, the primary controlled variable is head-nodding behavior. With IVE, all the variated behaviors is filtered out and the changes with head nods is presented. These head nods are highly likely to be disconnected with the virtual speaker's speech context and body movements. This is also the case with the original "Digital Chameleons" study. In the original "Digital Chameleons" study, the virtual characters have no other body movements except for the head movements. Although this setting is a bit unnatural, it avoided the disconnection between the virtual speaker's body movements and its head movements. However, the disconnection between the virtual speaker's speech context and its head movements still exists.

On the contrary, the experimental controllability has been made used fairly in the second experiment. For example, the number of speaker's head nods decreases in the dialogue compared to those in the monologue while the coordination of head nods between speaker and listener in dialogue (mimic, playback, natural conditions) is significantly higher than that in the monologue (recording condition). This further provides evidence for the collaborative dialogue theory.

7.4.2 Replicating the "Digital Chameleons" Paradigm

The "Digital Chameleons" paradigm refers to an agent automatically mimicking the other person's behavior and gaining social influence toward that person. This paradigm has been replicated in many research and with many alterations. For example, this paradigm was tested with different entities, including IVAs (Bailenson & Yee, 2005), mechanical hands (Bailenson et al., 2007), robots (Riek et al., 2010); with different tasks, including persuading (Bailenson & Yee, 2005), route planning and investment (Verberne et al., 2013), photo description (Hale & Hamilton, 2016); with different measures, including effectiveness (Bailenson & Yee, 2005), trust (Verberne et al., 2015; Verberne et al., 2013), likability (Stevens et al., 2016), rapport (Hale & Hamilton, 2016; Riek et al., 2010). Some of them succeed, some failed. The most of the original "Digital Chameleons" study is replicated in this study, e.g., the same task, the same measures and the same entities – IVAs are used. On the other hand, a more realistic IVE, better full-body motion capture system are used in this study. However, the result – mimicking IVAs were more effective than non-mimicking IVAs is failed to be replicated. This is not a surprise as it is explained in the previous sections.

The more important part for replicating the study is to collect head nods data with the experiments. However, it could be improved if implementing other tasks, for example, free communication.

7.4.3 Evaluating the Effectiveness of Virtual Speakers

In the study, the same measurements that the "Digital Chameleons" study used for the evaluation of the effectiveness of the IVA speakers as well as the avatar speakers are adopted. In this way, it could be ensured that it is not because of the use of a different measurement that the results are not consistent with the original study. On the other hand, the PCA method is used to create a less arbitrary composition for the components of the measurements. The PCA method turned out to be very useful in finding the significant difference in 'authority' between the recording condition and the mimic and playback conditions.

7.4.4 Behavior Analysis Methods

In the thesis, the analysis of head nods with different methods are described, including the linear statistical methods, frequency analysis, and nonlinear dynamic analysis (xRQA). These methods illustrated head nods from different points of view.

7.4.4.1 Linear Statistical Methods

The linear statistical methods used in this study include the descriptive statistics, Student T-test, ANOVA, and GLMM. Apart from the descriptive statistics, the Student T-test, ANOVA, and GLMM methods are used throughout the whole study as the group comparison methods after the preprocessing with other methods. These linear statistical methods are always useful since the nonlinear data can be always converted to the linear data with specific data transform techniques. Here, the descriptive statistics is discussed only.

The descriptive statistics in this study are the mean, max, standard deviation, and total head nods. The mean of head nods is not appealing since it would always be near 0. The max of head nods suggests the upper movement range of head nods in degree. The standard deviation of head nods indicates the range of degrees that most of the head nods lie in. The total head nods is the overall journey the head traveled in the head pitching direction.

These statistics are used to reveal some basic facts of the head nods in this study. For example, in the first experiment, 95% of listeners' head nods are deviating from the straight-ahead pose for no more than 10 degrees (2 $*$ standard deviation of head nods). In the second experiment, 95% of listeners' head nods are within 5 degrees instead.

These statistics are used to compare head nods between speakers, between listeners, and between speakers and listeners across conditions. These statistics cannot be used as the general facts of head nods since they are only applicable in this study for specific tasks. On the other hand, the comparison of these statistics between conditions would be useful. They could tell if there is any difference in speakers and listeners' head nods with different head nods manipulations. For example, The standard deviation of head nods and total head nods suggest no significant difference between listeners across conditions while speakers in the monologue (recording conditions) do significantly more head nods in the dialogue (mimic, playback, natural conditions).

7.4.4.2 Frequency Analysis

Frequency analysis on head nods in this study is inspired by the work of Hale et al. (2019). Hale et al. used Wavelet Analysis for the frequency analysis and suggested listeners nod more in the high-frequency domain (2.6-6.5 Hz) than speakers. However, performing Wavelet Analysis is quite complicated. It requires a systematic understanding of time-frequency transformation and a variety of kinds of wavelets. The product of Wavelet Analysis is a matrix of the intensity of a range of frequencies at a range of time points for the specific time-series. In terms of head nods time-series, the intensity of a specific frequency is not exactly the number of head nods at that frequency. Thus, it is not accurate to say listeners nod more in the high-frequency domain than speakers with Wavelet Analysis.

In this thesis, a more straightforward but still robust method is used for frequency analysis. It counts the exact amount of head nods by counting the peaks in the head nods time-series and extracts the number of head nods at a specific frequency with a low pass filter of which cutoff frequency slowly increasing. Then a simple paired T-tests is used to compare the number of head nods between speakers and listeners for each frequency band. This method used in this study on head nods time series suggests the same result as the work of Hale et al. (2019). That is, listeners nod more in the high-frequency domain (> 3 Hz) than speakers. Furthermore, it suggests that speakers nod more in the low-frequency domain (0.8-1.5 Hz) than listeners, which is consistent with the result from Healey et al. (2013). This suggests our method is more straightforward but as effective as Wavelet Analysis.

7.4.4.3 Cross Recurrence Quantification Analysis

For the measurement of coordination of head nods, the Cross Recurrence Quantification Analysis (xRQA) is used. xRQA takes two time-series and measures how one time-series is repeated by another or how similar are the two time-series.

xRQA produces Recurrent Plot, which is a powerful visualization for the similarity of two time-series over time. For example, Figure 6.6, Figure 6.12, and Figure 6.11 show us the different levels of coordination of head nods between speakers and listeners across conditions within a glance. xRQA further produces quantification outputs that are convenient for us to do further statistics and test whether there is a difference at a significant level. For example, the GLMM and ANOVA are used to compare the different levels of coordination of head nods and significant differences are found across conditions (Table 6.13 6.21).

On the other hand, xRQA could be challenging to use. Before performing xRQA, the parameters (delay, embedding dimension, radius, normalization) should first be found for it. Although there are automatic tools for the search of parameters – for example, Average Mutual Information function for 'delay', False Nearest Neighbor function for "embedding dimension" – it would still be difficult to decide what parameters to use when coming across xRQA for multiple groups of time-series. Perhaps the best practice is to use a set of fixed parameters for all the groups and settle the parameters so that no %REC and %DET output is restricted by the floor or ceiling effect, for example, the %REC should between 0.5% and 5%.

7.4.5 Conclusion

In this section, the methodological approaches, including the use of IVE, the replication of the "Digital Chameleons" paradigm, the evaluation of the effectiveness of speakers, and behavior analysis methods are discussed. The validity and limitations of these methods are investigated.

The results suggest that using IVE is promising for behavior study. However, it should be carefully designed at the same time. Replicating the "Digital Chameleons" paradigm is an excellent way to collect data. Evaluating the effectiveness of speakers with the same survey used in the original "Digital Chameleons" study ensured the consistency of the result. Different behavior analysis methods described head nods from different perspectives.

7.5 Summary

This chapter put together the findings of the two experiments, compared and discussed them reflectively. In the discussion, it is suggested that 1) the limitations of the "Digital Chameleons" paradigm are, first, the nonconscious mimicry behavior is rare in daily interaction; second, the effect of mimicry is subtle and fragile; third, the manipulation of head nods violated the communicative functions of head nods and ignored the different roles for speakers and listeners' head nods 2) Head nods are collaborative behavior, they are asymmetric between speakers and listeners in dialogue – listeners nod more in high-frequency domain, and less in the low-frequency domain than speakers; The coordination of head nods between speakers and listeners is not higher than chance level, mimicking virtual speakers increases head nods coordination and IVA speakers using prerecorded animation decreases head nods coordination 3) the using of IVE, replication of the "Digital Chameleons" are well-conducted but can be improved; the behavior analysis methods used in this study reveals difference perspective for head nods.

Chapter 8

Conclusions and Further Work

This chapter recapitulates the major findings in relation to the research goals presented in Chapter 1, as well as the contributions of this thesis. Limitations are discussed, and potential future works are indicated. The subject of this thesis is the study of the coordination of head nods between speakers and listeners in natural social interaction. The thesis set out with the general research question – How do speakers' and listeners' head nods become coordinated in dialogue? Based on the literature review on nonverbal communication and behavioral mimicry, more specific research questions are developed to examine the effects of the mimicry of head nods on the speaker's effectiveness, the coordination of different head nods.

The two empirical experiments together addressed the primary research goals of this thesis. Both of them consistently suggest the null effect of the mimicry of head nods on the effectiveness of the speaker. Full-body motion capture and realistic IVE are used in the study to facilitate social interaction. Speaker's head nods behavior is manipulated within each study and listener's presence is varied between the two studies to investigate the natural head nods coordination. Parallel to the investigation of the research questions, the thesis also developed a detailed understanding of head nods and explored the behavior analysis methods for modeling head nods with time-series data.

8.1 Major Findings

There are three sections of the major findings in this thesis. Each of the results can be linked to one of the research goals described in Section 1.2.2.

1. **Developing a customized IVE for the replication of the "Digital Chameleons" study.**

   The thesis described the architecture design of the IVE for the study of human interaction. It uses realistic settings with an office environment and realistic avatars. It offers a real-time natural interface enabled with a fine-grained full-body motion capture system for the interaction with an avatar or intelligent virtual agent. It provides the ability to manipulate avatar/agent's behavior – head nods in this study – for the replication of the "Digital Chameleons" study. It uses the multithreading technique for the best performance of the system and records the motion data of its users for future analysis. The IVE provides great controllability for the experiment.

2. **Evaluating the "Digital Chameleons" paradigm.**

   The two experiments described in the thesis repeatedly evaluated the "Digital Chameleons" paradigm and found consistent results. That is, the effectiveness of speakers do not increase when the speakers automatically mimicking the listeners' head nods. This suggests the subtle and fragile effect of the "Digital Chameleons" paradigm. The limitations of this paradigm is discussed in Section 7.2. The results suggest the difference of head nods between speakers and listeners is critical and should not be ignored in the social interaction, e.g., "Digital Chameleons" situation.

3. **Examining the head nods coordination in the dialogue.**

   The results from the two experiments suggest that head nods are highly role-specific nonverbal behavior (Section 2.1.2). The inappropriate manipulation of head nods could lead to an unnatural situation. Head nods are collaborative behaviors in dialogue. With the presence of the listeners, the speakers nod less in the dialogue than the speakers in a monologue. Head nods can be classified into slow, ordinary, and fast head nods. Fast head nods might be unnoticeable, but they are contributing to synchrony. Speakers and listeners' head nods are asymmetric in the high and low frequency domain. Specifically, speakers nod more in the low-frequency domain and less in the high-frequency domain than listeners. The coordination of head nods in the natural dialogue is no more than the chance level of head nods coordination. Speakers' and listeners' head nods become coordinated by collaboratively adjusting their head nods spontaneously.

Apart from these, different behavior analysis methods are compared. The thesis proposed a novel frequency analysis method that is straightforward and effective. The sophisticated analysis of the head nods behaviors suggest the fundamental difference of the behaviors. Lastly, this study also provided a good example of using Cross Recurrence Quantification Analysis for behavior studies.

8.2 Limitations and Future Works

As discussed earlier in Chapter 7, there were some limitations due to the methodology and study design.

Research Scope

The research subject is the coordination of head nods in dialogue. The study investigated speakers and listeners' head nods in the aspects of maximum head nods, total head nods, head nods in the specific frequency domain, and the similarity of head nods. However, the focus is limited. Apart from these statistics of head nods, the relation of particular head nods pattern with its function (Hadar et al., 1985) needs to be investigated in the future. The research of head nods is in the broader research field of nonverbal behavior (Mehrabian, 2017), which consist of gestures, postures, proxemics. As the full-body movements are collected in the experiments, it would be worth to take a look at these aspects in the future.

Human interactive systems are complex systems. It is essential to analyze human interaction with nonlinear methods. The study has taken the advantage of nonlinear analysis on head nods coordination with the cross recurrence quantification analysis (xRQA). It would also be interesting to analyze human head nods with other nonlinear methods and comparing these methods. Furthermore, as a lot of human behavior data is collected, some data-driven approaches, such as machine learning, can be used for behavior modeling.

This research used IVAs and avatars as speakers, which is partly in the research scope of the embodied conversational agent (ECA). However, the automatically mimicking IVA model is too simple that it did not contribute anything technically in the ECA field. Instead, some criticisms or 'cautions' are made to the development of ECAs. That is, if the ECA is aiming to deliver a natural interaction with someone, it would need context-aware ability, for example, taking account of the listener's head nods.

Immersive Virtual Environment technology is used in this study as the research tool for its ecology validity and experimental controllability. Apart from the integration of IVE with full-body motion capture, the study does not investigate more in the HCI part of virtual reality, for example, multimodal interaction in virtual reality (Makransky, Lilleholt, & Aaby, 2017), the haptic interface for virtual reality (Azmandian, Hancock, Benko, Ofek, & Wilson, 2016). There would be tremendous opportunities in these research fields, as well.

Design of IVE

As mentioned in Section 7.4.1.1, the design of IVE did not utilize the full potential of virtual reality. For example, a better virtual environment could be built to fit the tracking area in the laboratory. The motion tracking system could be improved with additional hand tracking (Scheggi, Meli, Pacchierotti, & Prattichizzo, 2015) and facial expression (Li et al., 2015) tracking. These could be done in future works.

Study Design

In the research, the "Digital Chameleons" study is replicated in both task design and condition design. In the study, the speaker is asked to deliver the university regulation of the student ID card to the listener. This is a two minutes short speech. The listener only needs to feedback with 'backchannels', e,g, head nods, "uh-huh". This task is quite limited.

Speaker and listener are not doing free conversation. A better task would be a photo description task (Hale & Hamilton, 2016), BART (Lejuez et al., 2002), etc. There are also concerns about these tasks, such as lack of control, difficult to analyze. Future research should decide what task to do after careful consideration.

The other limitation of the study design is the condition (experimental and controlled group) design. In the first experiment, there are 3 conditions 1) mimic – IVA mimic participant's head nods 2) playback – IVA using the previous participant's head nods 3) recording – IVA using a prerecorded head nods. The study complements the conditions in the "Digital Chameleons" study by adding the third recording condition, this is discussed in Section 7.2. In the second experiment, there are had 4 conditions 1) mimic – virtual speaker mimic participant's head nods 2) playback – virtual speaker using the previous participant's head nods 3) natural – virtual speaker using the real speaker's head nods 4) recording – virtual speaker using a prerecorded head nods. In this experiment, the natural condition is added in which real-time real-world speaker and listener's head nods are used. However, it still lacks one baseline situation where the speaker and listener converse without the intervention of the virtual reality. For future work, there could be the speaker and listener doing a natural conversation without using virtual reality.

Data Collection and Analysis

In this study, the same questionnaires are used as the one in the "Digital Chameleons" for the measurement of the effectiveness of the speakers. As noted in Section 7.4, this ensures that the inconsistent results are not coming from the use of different measurements . Instead, they might be resulted from cultural differences (US versus UK), era difference (2005 versus 2020), the educational difference (Stanford University versus QMUL). For the future, the questionnaires should be carefully redesign based on the current situation before the study.

Although videos are also recorded in this study, they are not used in the analysis in this thesis. They could help to analyze the relation of the nonverbal behavior and the semantics of the dialogue.

8.3 Closing Remarks

In the last month of the thesis writing, the world was suffering from the outbreak of the coronavirus COVID2019. People started to work from home and communicated with their colleagues using mobile phones or video conference tools. Jobs were getting harder because of inefficient communication. Life was getting boring because of being trapped in the room for too long. All these inconveniences are the potent reminders of the potentials of the IVEs. In the near future, people would be able to use VR in their daily work to communicate with each other efficiently with the rich spatial information and nonverbal cues, as the substitution of the video conference tools of today. People would be able to travel around in the IVE and get out of their homes, which detained them for a long time. But at the moment, there is still a lot of work to do to deliver the optimized VR experience to everyone. This thesis illustrated the experiments with IVE for behavior research. In addition to the better understanding of head nods coordination, I hope this thesis could push the VR communication application research a little bit further. Thus, people suffered from disease, poverty, and war could feel released when they put on their VR devices.

Bibliography

Ahn, S. J. (, Le, A. M. T., & Bailenson, J. (2013). The Effect of Embodied Experiences on Self-Other Merging, Attitude, and Helping Behavior. *Media Psychology*, (785022367).

Ahn, S. J., Bailenson, J. N., & Park, D. (2014). Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. *Computers in Human Behavior*, *39*(February), 235–245.

Allwood, J., & Cerrato, L. (2003). A study of gestural feedback expressions. In *First nordic symposium on multimodal communication* (pp. 7–22). Copenhagen.

Andersen, P. A. (1999). *Nonverbal communication: Forms and functions.* Mayfield Mountain View, CA.

Argelaguet, F. (2016). The Role of Interaction in Virtual Embodiment : Effects of the Virtual Hand Representation The Role of Interaction in Virtual Embodiment : Effects of the Virtual Hand Representation.

Ashton-James, C. E., & Chartrand, T. L. (2009). Social cues for creativity: The impact of behavioral mimicry on convergent and divergent thinking. *Journal of Experimental Social Psychology*, *45*(4), 1036–1040.

Azmandian, M., Hancock, M., Benko, H., Ofek, E., & Wilson, A. D. (2016). Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 1968–1979).

Bailenson, J. N. (2006). Transformed social interaction in collaborative virtual environments. In *Digital media: Transformations in human communication* (pp. 255–264).

Bailenson, J. N., Beall, A. C., & Blascovich, J. (2002). Gaze and task performance in shared virtual environments. *Journal of Visualization and Computer Animation*, *13*(5), 313–320.

Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., & Turk, M. (2004). Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, *13*(4), 428–441.

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

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal Distances in Virtual Environments. *Personality and Social Psychology Bulletin*, *29*(7), 819–833.

Bailenson, J. N., & Yee, N. (2003). The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context. *Journal of the Learning Sciences*, *12*(3), 361–407.

Bailenson, J. N., & Yee, N. (2005). Digital chameleons. *Society*, *16*(10), 814–819.

Bailenson, J. N., Yee, N., Patel, K., & Beall, A. C. (2008). Detecting digital chameleons. *Computers in Human Behavior*, *24*(1), 66–87.

Bailenson, J. N., Beall, a., Blascovich, J., Raimundo, M., & Weisbuch, M. (2001). Intelligent agents who wear your face: Users' reactions to the virtual self. *Intelligent Virtual Agents*, *2190/2001*, 86–99.

Bailenson, J. N., Yee, N., Brave, S., Merget, D., & Koslow, D. (2007). Virtual interpersonal touch: expressing and recognizing emotions through haptic devices. *Human-Computer Interaction*, *22*(November), 325–353.

Barber, P., & O'Leary, M. (1997). The relevance of salience: Towards an activational account of irrelevant stimulus-response compatibility effects. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus-response compatibility* (Vol. 118, pp. 135–172). Advances in Psychology. North-Holland.

Bardi, A., & Zentner, M. (2017). Grand challenges for personality and social psychology: Moving beyond the replication crisis. *Frontiers in psychology*, *8*, 2068.

Bargh, J. A., & Shalev, I. (2012). The substitutability of physical and social warmth in daily life. *Emotion*, *12*(1), 154.

Battersby, S. A., & Healey, P. G. T. (2010a). Using head movement to detect listener responses during multi-party dialogue. In *Proceedings of lrec workshop on multimodal corpora advances in capturing coding and analysing multimodality* (pp. 11–15).

Battersby, S. A., & Healey, P. G. T. (2010b). Head and Hand Movements in the Orchestration of Dialogue. *Proceedings of the Cognitive Science Society*, *32*(32).

Bavelas, J. B., Black, A., Lemery, C. R., & Mullett, J. (1986). " i show how you feel": Motor mimicry as a communicative act. *Journal of personality and social psychology*, *50*(2), 322.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of personality and social psychology*, *79*(6), 941.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, *52*(3), 566–580.

Bavelas, J. B., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, *13*(1), 63–92.

Beall, A. C., Bailenson, J. N., Loomis, J., Blascovich, J., & Rex, C. S. (2003). Non-zero-sum gaze in immersive virtual environments. *HCI International*.

Bee, N., André, E., & Tober, S. (2009). Breaking the ice in human-agent communication: Eye-gaze based initiation of contact with an embodied conversational agent. In Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Eds.), *Intelligent virtual agents* (pp. 229–242). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bell, A. (1984). Language style as audience design. *Language in society*, *13*(2), 145–204.

Bergmann, K., Branigan, H. P., & Kopp, S. (2015). Exploring the alignment space – lexical and gestural alignment with real and virtual humans. *Frontiers in ICT*, *2*, 7.

Bergström, I., Kilteni, K., & Slater, M. (2016). First-person perspective virtual body posture influences stress: A virtual reality body ownership study. *PloS one*, *11*(2), e0148060.

Berkowitz, L. (1984). Some effects of thoughts on anti-and prosocial influences of media events: A cognitive-neoassociation analysis. *Psychological bulletin*, *95*(3), 410.

Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal behavior*, *12*(2), 120–138.

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony.

Bernieri, F., Reznick, J., & Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology*, *54*, 243–253.

Beskow, J., Peters, C., Castellano, G., O'Sullivan, C., Leite, I., & Kopp, S. (2017). Intelligent virtual agents. In *17th international conference. proceedings. springer lncs* (Vol. 10498). Springer.

Bhagat, K. K., Liou, W.-K., & Chang, C.-Y. (2016). A cost-effective interactive 3d virtual reality system applied to military live firing training. *Virtual Reality*, *20*(2), 127–140.

Biocca, F., & Levy, M. R. (2013). *Communication in the age of virtual reality*. Routledge.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124.

Bohbot, V. D., Lerch, J., Thorndycraft, B., Iaria, G., & Zijdenbos, A. P. (2007). Gray matter differences correlate with spontaneous strategies in a human virtual navigation task. *Journal of Neuroscience*, *27*(38), 10078–10083.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature reviews neuroscience*, *12*(12), 752.

Boholm, M., & Allwood, J. (2010). Repeated head movements, their function and relation to speech. In *Proceedings of the seventh conference on international language resources and evaluation (lrec'10)* (pp. 6–10). Citeseer.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2019). Studying human behavior with virtual reality: The unity experiment framework. *Behavior Research Methods*.

Buller, D. B., & Aune, R. K. (1988). The effects of vocalics and nonverbal sensitivity on compliance: A speech accommodation theory explanation. *Human Communication Research*, *14*(3), 301–332.

Burgoon, J. K., Birk, T., & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human communication research*, *17*(1), 140–169.

Cappella, J. N. (2005). Coding mutual adaptation in dyadic nonverbal interaction. *The sourcebook of nonverbal measures: Going beyond words*, 383–392.

Carpenter, W. B. (1875). *Principles of mental physiology..* Henry S. King&Company.

Carver, C. S., Ganellen, R. J., Froming, W. J., & Chambers, W. (1983). Modeling: An analysis in terms of category accessibility. *Journal of Experimental Social Psychology*, *19*(5), 403–421.

Cassell, J., Sullivan, J., Churchill, E., & Prevost, S. (2000). *Embodied conversational agents*.

Castelli, L., Pavan, G., Ferrari, E., & Kashima, Y. (2009). The stereotyper and the chameleon: The effects of stereotype use on perceivers' mimicry. *Journal of Experimental Social Psychology*, *45*(4), 835–839.

Catmur, C., & Heyes, C. (2013). Is It What You Do, or When You Do It? The Roles of Contingency and Similarity in Pro-Social Effects of Imitation. *Cognitive Science*, *37*(8), 1541–1552.

Charny, E. J. (1969). Psychosomatic manifestations of rapport in psychotherapy. *General Systems Theory and Psychiatry*, 267.

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.

Chartrand, T. L., & Dalton, A. N. (2009). Mimicry: Its ubiquity, importance, and functionality. *Oxford handbook of human action*, 458–483.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual review of psychology*, *64*, 285–308.

Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. *The new unconscious*, 334–361.

Chartrand, T. L., & Van Baaren, R. (2009). Human mimicry. *Advances in experimental social psychology*, *41*, 219–274.

Clark, A. E., & Kashima, Y. (2007). Stereotypes help people connect with others in the community: A situated functional analysis of the stereotype consistency bias in communication. *Journal of personality and social psychology*, *93*(6), 1028.

Clark, H. H. (1996). *Using Language*.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. L. Ny & W. Kintsch (Eds.), *Language and comprehension* (Vol. 9, pp. 287–299). Advances in Psychology. North-Holland.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*(2), 332–373.

Condon, W. S., & Ogston, W. D. (1967). A segmentation of behavior. *Journal of psychiatric research*.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences*, *37*(2), 177–192.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The cave: Audio visual experience automatic virtual environment. *Communications of the ACM*, *35*(6), 64–73.

Daher, S., Kim, K., Lee, M., Schubert, R., Bruder, G., Bailenson, J., & Welch, G. (2017). Effects of social priming on social presence with intelligent virtual agents. In J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, & S. Kopp (Eds.), *Intelligent virtual agents* (pp. 87–100). Cham: Springer International Publishing.

Dale, R., & Spivey, M. J. (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.

Dalton, A. N., Chartrand, T. L., & Finkel, E. J. (2010). The schema-driven chameleon: How mimicry affects executive and self-regulatory resources. *Journal of Personality and social Psychology*, *98*(4), 605.

Davis, J. W., & Bobick, A. F. (1997). The representation and recognition of human movement using temporal templates. In *Proceedings of ieee computer society conference on computer vision and pattern recognition* (pp. 928–934). IEEE.

Davis, M. H. et al. (1980). A multidimensional approach to individual differences in empathy.

De La Rosa, S., Ferstl, Y., & Bülthoff, H. H. (2016). Visual adaptation dominates bimodal visual-motor action adaptation. *Scientific reports*, *6*, 23829.

De Leo, G., Diggs, L. A., Radici, E., & Mastaglio, T. W. (2014). Measuring sense of presence and user characteristics to predict effective training in an online simulated virtual environment. *Simulation in Healthcare*, *9*(1), 1–6.

de Melo, C. M., Gratch, J., & Carnevale, P. J. (2015). Humans versus computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing*, *6*(2), 127–136.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, *3*(3), 349–365.

Dennett, D. C., & Dennett, D. C. (1996). *Darwin's dangerous idea: Evolution and the meanins of life*. Simon and Schuster.

Deryugina, O. (2010). Chatterbots. *Scientific and Technical Information Processing*, *37*(2), 143–147.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180.

Diener, E., & Biswas-Diener, R. (2016). The replication crisis in psychology.

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In *Advances in experimental social psychology* (Vol. 33, pp. 1–40). Elsevier.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*(1), 86–89. PMID: 11228851.

Dinstein, I., Thomas, C., Behrmann, M., & Heeger, D. J. (2008). A mirror up to nature. *Current Biology*, *18*(1), R13–R18.

Dodds, T. J., Mohler, B. J., & Bülthoff, H. H. (2011). Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PLoS ONE*, *6*(10).

Driscoll, I., Hamilton, D. A., Yeo, R. A., Brooks, W. M., & Sutherland, R. J. (2005). Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Hormones and Behavior*, *47*(3), 326–335.

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in psychology*, *6*, 621.

Edinger, J. A., & Patterson, M. L. (1983). Nonverbal involvement and social control. *Psychological Bulletin*, *93*(1), 30.

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, *317*(5841), 1048–1048.

Evans, D. (2002). *Emotion: The science of sentiment*. Oxford University Press, USA.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611.

Fallis, A. (2013). Mutual gaze and personal space in virtual environments. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. eprint: arXiv:1011.1669v3

Finkel, E. J., Campbell, W. K., Brunell, A. B., Dalton, A. N., Scarbeck, S. J., & Chartrand, T. L. (2006). High-maintenance interaction: Inefficient social coordination impairs self-regulation. *Journal of personality and social psychology*, *91*(3), 456.

Fischer-Lokou, J., Martin, A., Guéguen, N., & Lamy, L. (2011). Mimicry and propagation of prosocial behavior in a natural setting. *Psychological Reports*, *108*(2), 599–605. PMID: 21675573.

Fox, J., & Bailenson, J. N. (2009). Virtual virgins and vamps: The effects of exposure to female characters' sexualized appearance and gaze in an immersive virtual environment. *Sex Roles*, *61*(3-4), 147–157. arXiv: arXiv:1011.1669v3

Fox, J., Fox, J., & Hall, D. (n.d.). Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence Jesse. (614).

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of personality and Social Psychology*, *62*(3), 378.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609.

Garau, M., Slater, M., Pertaub, D.-P., & Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, *14*(1), 104–116.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 529–536). CHI '03. Ft. Lauderdale, Florida, USA: Association for Computing Machinery.

Garfinkel, H. (1964). Studies of the routine grounds of everyday activities. *Social problems*, *11*(3), 225–250.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, *16*(18), 1824–1829.

Gerwing, J., & Bavelas, J. B. (2013). The social interactive nature of gestures - Theory, assumptions, methods, and findings. In *Body - language - communication* (pp. 821–836).

Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo*, *11*(3), 227–235.

Giles, H. E., Coupland, J. E., & Coupland, N. E. (1991). Contexts of accommodation: Developments in applied sociolinguistics.

Giorgolo, G., & Verstraten, F. A. J. (2008). Perception of speech-and-gesture integration. In *In proceedings of the international conference on auditory-visual speech processing* (pp. 31–36).

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual review of psychology*, *64*, 257–283.

Goodwin, C. (1979). The Interactive Construction of a Sentence in Natural Conversation, 97–121.

Goodwin, M. H. (1985). The Serious Side of Jump Rope: Conversational Practices and Social Organization in the Frame of Play. *The Journal of American Folklore*, *98*(389), 315.

Grammer, K., Honda, M., Juette, A., & Schmitt, A. (1999). Fuzziness of nonverbal courtship communication unblurred by motion energy detection. *Journal of personality and social psychology*, *77*(3), 487.

Grassberger, P., Schreiber, T., & Schaffrath, C. (1991). Nonlinear time sequence analysis. *International Journal of Bifurcation and Chaos*, *1*(03), 521–547.

Gray, A. (2016). Cross recurrence quantification analysis tutorial.

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological review*, *77*(2), 73.

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & McCall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, *10*(1), 1–22.

Gueguen, N. (2011). The mimicker is a mirror of myself: Impact of mimicking on self-consciousness and social anxiety. *Social Behavior and Personality: an international journal*, *39*(6), 725–728.

Guerrero, L. K., & Floyd, K. (2006). *Nonverbal communication in close relationships*. Routledge.

Gurion, T., Healey, P., & Hough, J. (2018). Real-time testing of nonverbal interaction: An experimental method and platform.

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

Hadar, U., Steiner, T. J., & Rose, F. C. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, *9*(4), 214–228.

Hale, J., & Hamilton, A. F. D. C. (2016). Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific Reports*, *6*.

Hale, J., Payne, M. E., Taylor, K. M., Paoletti, D., & Hamilton, A. F. D. C. (2018). The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology*, *71*(4), 989–1008.

Hale, J., Ward, J. A., Buccheri, F., Oliver, D., & Hamilton, A. F. d. C. (2019). Are you on my wavelength? interpersonal coordination in dyadic conversations. *Journal of Nonverbal Behavior*, 1–21.

Hall, E. T., Birdwhistell, R. L., Bock, B., Bohannan, P., Diebold Jr, A. R., Durbin, M., ... Kimball, S. T., et al. (1968). Proxemics [and comments and replies]. *Current anthropology*, *9*(2/3), 83–108.

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice*. Routledge.

Hantula, D. A. (2019). Replication and reliability in behavior science and behavior analysis: A call for a conversation. Springer.

Harakeh, Z., Engels, R. C., Baaren, R. B. V., & Scholte, R. H. (2007). Imitation of cigarette smoking: An experimental study on smoking in a naturalistic setting. *Drug and Alcohol Dependence*, *86*(2), 199–206.

Hargie, O. (2016). *Skilled interpersonal communication: Research, theory and practice 6th edition*.

Harrist, A., & Waugh, R. (2013). Dyadic synchrony: Its structure and function in children's development. *Developmental Review*, 555–592.

Hasler, B. S., Hirschberger, G., Shani-Sherman, T., & Friedman, D. A. (2014). Virtual peacemakers: Mimicry increases empathy in simulated contact with virtual outgroup members. *Cyberpsychology, Behavior, and Social Networking*, *17*(12), 766–771.

Hatfield, E., Rapson, R., & Le, Y. (2009). Primitive emotional contagion: Recent research. *The social neuroscience of empathy*, 19–30.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–100.

Hawk, S. T., Fischer, A. H., & Van Kleef, G. A. (2011). Taking your place or matching your face: Two paths to empathic embarrassment. *Emotion*, *11*(3), 502.

Healey, P. G. T., Lavelle, M., Howes, C., Battersby, S., & McCabe, R. (2013). How listeners respond to speaker's troubles. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35, *35*).

Healey, P. G. T., Plant, N., Howes, C., & Lavelle, M. (2015). When words fail: Collaborative gestures during clarification dialogues. In *2015 aaai spring symposium series,(chicago)*.

Healey, P. G. T., Frauenberger, C., Gillies, M., & Battersby, S. (2009). Experimenting with Non-Verbal Interaction, 2–3.

Helt, M. S., Eigsti, I.-M., Snyder, P. J., & Fein, D. A. (2010). Contagious yawning in autistic and typical development. *Child development*, *81*(5), 1620–1631.

Heritage, J., & Atkinson, J. M. (1984). Structures of social action. *Studies in Conversation Analysis*, 346–369.

Hermans, R. C., Lichtwarck-Aschoff, A., Bevelander, K. E., Herman, C. P., Larsen, J. K., & Engels, R. C. (2012). Mimicry of food intake: The dynamic interplay between eating companions. *PloS one*, *7*(2), e31027.

Herrmann, A., Rossberg, N., Huber, F., Landwehr, J. R., & Henkel, S. (2011). The impact of mimicry on sales – evidence from field and lab experiments. *Journal of Economic Psychology*, *32*(3), 502–514.

Heslin, R., & Alper, T. (1983). Touch: A bonding gesture. *Nonverbal interaction*, 47–75.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, *34*(4), 575–583.

Heyes, C. (2011). Automatic imitation. *Psychological bulletin*, *137*(3), 463.

Heylen, D. (2005). Challenges ahead: Head movements and other social acts in conversations. In *Proceedings of joint symposium on virtual social agents* (pp. 45–52).

Heylen, D. (2006). Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics*, *3*(03), 241–267.

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods*, *49*(1), 46–60.

Hickson, M., Stacks, D. W., & Moore, N.-J. (2004). *Nonverbal communication: Studies and applications*. Roxbury Pub.

Hinde, R. A. (1972). *Non-verbal communication*. Cambridge University Press.

Holler, J., & Wilkin, K. (2011a). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, *35*(2), 133–153.

Holler, J., & Wilkin, K. (2011b). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, *35*(2), 133–153.

Hove, M. J., & Risen, J. L. (2009). It's All in the Timing: Interpersonal Synchrony Increases Affiliation. *Social Cognition*, *27*(6), 949–960.

Huang, L., Morency, L.-P., & Gratch, J. (2011). Virtual rapport 2.0. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, & K. R. Thórisson (Eds.), *Intelligent virtual agents* (pp. 68–79). Berlin, Heidelberg: Springer Berlin Heidelberg.

Huntsinger, J. R., Lun, J., Sinclair, S., & Clore, G. L. (2009). Contagion without contact: Anticipatory mood matching in response to affiliative motivation. *Personality and Social Psychology Bulletin*, *35*(7), 909–922. PMID: 19487484.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526–2528.

Inzlicht, M., Gutsell, J. N., & Legault, L. (2012). Mimicry reduces racial prejudice. *Journal of Experimental Social Psychology*, *48*(1), 361–365.

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of personality and social psychology*, *99*(3), 549.

Ishi, C. T., Ishiguro, H., & Hagita, N. (2014a). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, *57*, 233–243.

Ishi, C. T., Ishiguro, H., & Hagita, N. (2014b). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, *57*, 233–243.

Iva '19: Proceedings of the 19th acm international conference on intelligent virtual agents. (2019), Paris, France: Association for Computing Machinery.

Jacob, C., Guéguen, N., Martin, A., & Boulbry, G. (2011). Retail salespeople's mimicry of customers: Effects on consumer behavior. *Journal of Retailing and Consumer Services*, *18*(5), 381–388.

James, W., Burkhardt, F., Bowers, F., & Skrupskelis, I. K. (1890). *The principles of psychology*. Macmillan London.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–202.

Jeannerod, M., & Jeannerod, M. (1997). *The cognitive neuroscience of action*. Blackwell Oxford.

Johnston, L. (2002). Behavioral mimicry and stigmatization. *Social Cognition*, *20*(1), 18–35.

Jones, S. E. (1999). Communicating with touch. *V The Nonverbal Communication Reader: Classic and contemporary readings, ur. Laura K. Guerrero, Joseph A. DeVito in Michael L. Hecht*, 192–201.

Keller, P., Knoblich, G., & Repp, B. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and cognition*, *16*, 102–11.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. CUP Archive.

Kennel, M. B., Brown, R., & Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, *45*(6), 3403.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Keysers, C., & Gazzola, V. (2010). Social neuroscience: Mirror neurons recorded in humans. *Current Biology*, *20*(8), R353–R354.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior, v.31, 354-364 (2010)*, *31*.

Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science - CURR DIRECTIONS PSYCHOL SCI*, *15*, 99–104.

Kothgassner, O. D., Griesinger, M., Kettner, K., Wayan, K., V??lkl-Kernstock, S., Hlavacs, H., … Felnhofer, A. (2017). Real-life prosocial behavior decreases after being socially excluded by avatars, not agents. *Computers in Human Behavior*, *70*, 261–269.

Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology*, *6*(2), 81–98.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In *Advances in experimental social psychology* (Vol. 28, pp. 389–450). Elsevier.

Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Journal of personality and social psychology*, *43*(4), 718.

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, *47*(7), 567–587.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72.

Kuhlen, A. K. (2010). *Assessing and accommodating addressees' needs: The role of speakers' prior expectations and addressees' feedback* (Doctoral dissertation, The Graduate School, Stony Brook University: Stony Brook, NY.).

La France, M., & Ickes, W. (1981). Posture mirroring and interactional involvement: Sex and sex typing effects. *Journal of Nonverbal behavior*, *5*(3), 139–154.

LaFrance, M. (1979). Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Social Psychology Quarterly*, 66–70.

LaFrance, M. (1982). Posture mirroring and rapport. *Interaction rhythms: Periodicity in communicative behavior*, *279298*.

Lafrance, M., & Broadbent, M. (1976). Group rapport: Posture sharing as a nonverbal indicator. *Group & Organization Studies*, *1*(3), 328–333.

Lakens, D. (2010). Movement synchrony and perceived entitativity. *Journal of Experimental Social Psychology*, *46*, 701–708.

Lakin, J. L. (2013). Behavioral mimicry and interpersonal synchrony. *Nonverbal communication*, 539–575.

Lakin, J. L., & Chartrand, T. L. (2003). Using Nonconscious Behavioral Mimicry to Create Affiliation and Rapport. *Psychological Science*, *14*(4), 334–339.

Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological Science*, *19*(8), 816–822. PMID: 18816290.

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd acm symposium on virtual reality software and technology* (p. 39). ACM.

Leander, N. P., Chartrand, T. L., & Bargh, J. A. (2012). You give me the chills: Embodied reactions to inappropriate amounts of behavioral mimicry. *Psychological Science*, *23*(7), 772–779. PMID: 22609538.

Lee, J., Wang, Z., & Marsella, S. (2010). Evaluating models of speaker head nods for virtual agents. In *Proceedings of the 9th international conference on autonomous agents and multiagent systems: Volume 1-volume 1* (pp. 1257–1264). International Foundation for Autonomous Agents and Multiagent Systems.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., … Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, *8*(2), 75.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, *317*(5841), 1096–1099.

Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*(1), 78–106.

Levenson, R. W., & Ruef, A. M. (1997). Physiological aspects of emotional knowledge and rapport.

Lhermitte, F. (1983). 'utilization behaviour'and its relation to lesions of the frontal lobes. *Brain*, *106*(2), 237–255.

Li, H., Trutoiu, L., Olszewski, K., Wei, L., Trutna, T., Hsieh, P.-L., … Ma, C. (2015). Facial performance sensing head-mounted display. *ACM Transactions on Graphics (ToG)*, *34*(4), 1–9.

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., & Weyers, P. (2008). Modulation of facial mimicry by attitudes. *Journal of Experimental Social Psychology*, *44*(4), 1065–1072.

Likowski, K. U., Weyers, P., Seibt, B., Stohr, C., Pauli, P., & Mhlberger, A. (2011). Sad and lonely? sad mood suppresses facial mimicry. *Journal of Nonverbal Behavior*, *35*(2), 101–117.

Lilienfeld, S. O. (2017). Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on psychological science*, *12*(4), 660–664.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic bulletin & review*, *9*(3), 550–557.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological review*, *91*(3), 295.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *"Behavior Research Methods, Instruments, & Computers"*, *31*(4), 557–564.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, *20*(2), 130–141.

Lundqvist, L.-O., & Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology*.

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285.

Marsh, K., Richardson, M., & Schmidt, R. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*, 320–339.

Martin, J. N., & Nakayama, T. K. (2013). *Intercultural communication in contexts*. McGraw-Hill New York, NY.

Maselli, A., & Slater, M. (2013). The building blocks of the full body ownership illusion. *Frontiers in human neuroscience*, *7*, 83.

Mast, M. S., & Cousin, G. (2013). Power, Dominance, and Persuasion. In J. A. Hall & M. L. Knapp (Eds.), *Nonverbal communication* *(handbooks of communication science, hocs 2)* (Chap. 20, pp. 613–635). Berlin, Boston: DE GRUYTER.

Maurer, R. E., & Tindall, J. H. (1983). Effect of postural congruence on client's perception of counselor empathy. *Journal of counseling psychology*, *30*(2), 158.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? what does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487.

Mazzurega, M., Pavani, F., Paladino, M., & Schubert, T. (2011). Self-other bodily merging in the context of synchronous but arbitrary-related multisensory inputs. *213*, 213–21.

McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological review*, *90*(3), 215.

McCall, C., & Singer, T. (2015). Facing off with unfair others: Introducing proxemic imaging as an implicit measure of approach and avoidance during social interaction. *PloS one*, *10*(2), e0117532.

McIntosh, D. N. (2006). Spontaneous facial mimicry, liking and emotional contagion. *Polish Psychological Bulletin*, *37*(1), 31.

Mehrabian, A. (2017). *Nonverbal communication*. Routledge.

Mehrabian, A., & Williams, M. (1969). Nonverbal concomitants of perceived and intended persuasiveness. *Journal of Personality and Social psychology*, *13*(1), 37.

Milner, D., & Goodale, M. (2006). *The visual brain in action*. OUP Oxford.

Mohler, B. J., Bülthoff, H. H., Thompson, W. B., & Creem-Regehr, S. H. (2008). A full-body avatar improves egocentric distance judgments in an immersive virtual environment. *APGV '08: Proceedings of the 5th symposium on Applied perception in graphics and visualization*, (August), 194.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fmri studies. *Neuroscience & Biobehavioral Reviews*, *36*(1), 341–349.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics Automation Magazine*, *19*(2), 98–100.

Navarro, J. (2008). *What every body is saying: An ex-FBI agent's guide to speed-reading people.* (M. Karlins, Ed.). New York, NY: Collins.

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: Amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673–678.

Néda, Z., Regan, E., Bréchet, Y., Vicsek, T., & Barabasi, A.-L. (2000). The sound of many hands clapping. *Nature*, *403*, 849–50.

Neumann, R., & Strack, F. (2000). " mood contagion": The automatic transfer of mood between persons. *Journal of personality and social psychology*, *79*(2), 211.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360.

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, *5*, 114.

Oppy, G., & Dowe, D. (2003). The turing test.

Oullier, O., Guzman, G., Jantzen, K., Lagarde, J., & Kelso, S. (2008). Social coordination dynamics: Measuring human bonding. *Social neuroscience*, *3*, 178–92.

Over, H., & Carpenter, M. (2009). Priming third-party ostracism increases affiliative imitation in children. *Developmental science*, *12*(3), F1–F8.

Paladino, M., Mazzurega, M., Pavani, F., & Schubert, T. (2010). Synchronous multisensory stimulation blurs self-other boundaries. *Psychological science*, *21*, 1202–7.

Pan, X., Gillies, M., Barker, C., Clark, D. M., & Slater, M. (2012). Socially anxious and confident men interact with a forward virtual woman: An experimental study. *PloS one*, *7*(4), e32931.

Pan, X., & Hamilton, A. F. d. C. (2015). Automatic imitation in a rich social context with virtual characters. *Frontiers in Psychology*, *6*, 790.

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417.

Pan, Y., & Steed, A. (2016). A comparison of avatar-, video-, and robot-mediated interaction on users' trust in expertise. *Frontiers in Robotics and AI*, *3*, 12.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*, 660.

Pascolo, P., & Budai, R. (2013). Just how consistent is the mirror neuron system paradigm. *Progress in Neuroscience*, *1*(1-4), 29–43.

Passingham, R. E. (1993). *The frontal lobes and voluntary action.* Oxford University Press.

Paus, T., Petrides, M., Evans, A. C., & Meyer, E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of neurophysiology*, *70*(2), 453–469.

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., & Nass', C. I. (2006). Adaptive language behavior in hci: How expectations and beliefs about a system affect users' word choice. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1177–1180). CHI '06. Montréal, Québec, Canada: Association for Computing Machinery.

Pease, B., & Pease, A. (2008). *The definitive book of body language: The hidden meaning behind people's gestures and expressions.* Bantam.

Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review*, *26*(3), 894–900.

Pejsa, T., Gleicher, M., & Mutlu, B. (2017). Who, me? how virtual agents can shape conversational footing in virtual reality. In J. Beskow, C. Peters, G. Castellano, C. O'Sullivan, I. Leite, & S. Kopp (Eds.), *Intelligent virtual agents* (pp. 347–359). Cham: Springer International Publishing.

Peña, J., Hancock, J. T., & Merola, N. A. (2009). The priming effects of avatars in virtual settings. *Communication Research*, *36*(6), 838–856.

Pickering, M. J., & Garrod, S. (2004a). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190.

Pickering, M. J., & Garrod, S. (2004b). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169–190.

Plant, N. J. (2018). *Intersubjectivity, empathy and nonverbal interaction* (Doctoral dissertation, Queen Mary University of London).

Pobric, G., & de C. Hamilton, A. F. (2006). Action understanding requires the left inferior frontal cortex. *Current Biology*, *16*(5), 524–529.

Prichard, D., & Theiler, J. (1994). Generalized redundancies for time series analysis. *arXiv preprint comp-gas/9405006*.

Priestley, M. B. (1988). Non-linear and non-stationary time series analysis.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167–201). Berlin, Heidelberg: Springer Berlin Heidelberg.

Provine, R. R. (1986). Yawning as a stereotyped action pattern and releasing stimulus. *Ethology*, *72*(2), 109–122.

Raij, A. B., Johnsen, K., Dickerson, R. F., Lok, B. C., Cohen, M. S., Duerson, M., ... Lind, D. S. (2007). Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE Transactions on Visualization and Computer Graphics*, *13*(3), 443–457.

Ramseyer, F., & Tschacher, W. (2006). Synchrony: A core concept for a constructivist approach to psychotherapy. *Constructivism in the Human Sciences*, *11*, 150–171.

Redeker, M., Stel, M., & Mastop, J. (2011). Does mimicking others change your self-view? *The Journal of Social Psychology*, *151*(4), 387–390. PMID: 21755649.

Reeves, S., Kuper, A., & Hodges, B. D. (2008). Qualitative research methodologies: Ethnography. *BMJ*, *337*.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, *18*(5), 407–413.

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

Richardson, M. J., Lopresti-Goodman, S., Mancini, M., Kay, B., & Schmidt, R. (2008). Comparing the attractor strength of intra-and interpersonal interlimb coordination using cross-recurrence analysis. *Neuroscience Letters*, *438*(3), 340–345.

Richardson, M., Marsh, K., Isenhower, R., Goodman, J., & Schmidt, R. (2008). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human movement science*, *26*, 867–91.

Riek, L. D., Paul, P. C., & Robinson, P. (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.

Riva, G., Raspelli, S., Algeri, D., Pallavicini, F., Gorini, A., Wiederhold, B. K., & Gaggioli, A. (2010). Interreality in practice: Bridging virtual and real worlds in the treatment of posttraumatic stress disorders. *Cyberpsychology, Behavior, and Social Networking*, *13*(1), 55–65.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169–192. PMID: 15217330.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews neuroscience*, *2*(9), 661.

Roberts, G. L. (1996). The communicative dictionary: A collaborative theory of meaning1 gillian l. roberts and janet beavin bavelas. *Beyond the symbol model: Reflections on the representational nature of language*, 135.

Romano, M. C., Thiel, M., Kurths, J., & von Bloh, W. (2004). Multivariate recurrence plots. *Physics letters A*, *330*(3-4), 214–223.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, *130*(9), 2452–2461.

Scheflen, A. E. (1964). The significance of posture in communication systems. *Psychiatry*, *27*(4), 316–331.

Scheggi, S., Meli, L., Pacchierotti, C., & Prattichizzo, D. (2015). Touch the virtual reality: Using the leap motion controller for hand track-

ing and wearable tactile devices for immersive haptic rendering. In *Acm siggraph 2015 posters* (pp. 1–1).
Schmidt, R., & Carello, C. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of experimental psychology. Human perception and performance*, *16*, 227–47.
Schmitz, A. (2012). A primer on communication studies. *Retrieved September*, *19*, 2016.
Schmuckler, M. A. (2001). What is ecological validity? a dimensional analysis. *Infancy*, *2*(4), 419–436.
Schober, M. F., Conrad, F. G., & Fricker, S. S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology*, *18*(2), 169–188.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379–423.
Shockley, K., Santana, M.-V., & Fowler, C. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of experimental psychology. Human perception and performance*, *29*, 326–32.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.
Skinner, B. (1938). The behavior of organisms: An experimental analysis.
Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557.
Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., … Sanchez-Vives, M. V. (2006). A virtual reprise of the stanley milgram obedience experiments. *PloS one*, *1*(1), e39.
Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PloS one*, *5*(5), e10564.
Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, *9*(5), 413–434.

Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.* *2*(3), 201–219.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*(5408), 1657–1661.

Stel, M., Blascovich, J., McCall, C., Mastop, J., Van Baaren, R. B., & Vonk, R. (2010). Mimicking disliked others: Effects of a priori liking on the mimicry-liking link. *European Journal of Social Psychology*, *40*(5), 867–880.

Stel, M., & Harinck, F. (2011). Being mimicked makes you a prosocial voter. *Experimental psychology.*

Stel, M., Mastop, J., & Strick, M. (2011). The impact of mimicking on attitudes toward products presented in tv commercials. *Social Influence*, *6*(3), 142–152.

Stel, M., Van Baaren, R. B., & Vonk, R. (2008). Effects of mimicking: Acting prosocially by being emotionally moved. *European journal of social psychology*, *38*(6), 965–976.

Stel, M., van Dijk, E., & Olivier, E. (2009). You want to know the truth? then don't mimic! *Psychological Science*, *20*(6), 693–699. PMID: 19422628.

Stel, M., & Vonk, R. (2010). Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction. *British Journal of Psychology*, *101*(2), 311–323.

Stevens, C. J., Pinchbeck, B., Lewis, T., Luerssen, M., Pfitzner, D., Powers, D. M., ... Gibert, G. (2016). Mimicry and expressiveness of an eca in human-agent interaction: Familiarity breeds content! *Computational cognitive science*, *2*(1), 1.

Swol, L. M. V., & Drury-Grogan, M. L. (2017). The effects of shared opinions on nonverbal mimicry. *SAGE Open*, *7*(2), 2158244017707243.

Takakura, I. T., Hoshi, R. A., Santos, M. A., Pivatelli, F. C., Nóbrega, J. H., Guedes, D. L., ... Godoy, M. F. d. (2017). Recurrence plots: A new tool for quantification of cardiac autonomic nervous system recovery after transplant. *Brazilian journal of cardiovascular surgery*, *32*(4), 245–252.

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, warwick 1980* (pp. 366–381). Springer.

Tanner, R. J., Ferraro, R., Chartrand, T. L., Bettman, J. R., & Baaren, R. V. (2007). Of Chameleons and Consumption: The Impact of Mimicry on Choice and Preferences. *Journal of Consumer Research*, *34*(6), 754–766.

Thornhill, R., & Gangestad, S. W. (1999). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and human behavior*, *20*(3), 175–201.

Tia, B., Saimpont, A., Paizis, C., Mourey, F., Fadiga, L., & Pozzo, T. (2011). Does observation of postural imbalance induce a postural reaction? *PloS one*, *6*(3), e17799.

Tickle-Degnen, L. (2006). Nonverbal behavior and its functions in the ecosystem of rapport.

Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of personality and social psychology*, *84*(3), 558.

Tincani, M., & Travers, J. (2019). Replication research, publication bias, and applied behavior analysis. *Perspectives on Behavior Science*, *42*(1), 59–75.

Toczek, Y. (2016). The Influence of Visual Realism on the Sense of Presence in Virtual Environments. (September).

Traiger, S. (2003). Making the right identification in the turing test. In *The turing test* (pp. 99–110). Springer.

Tranel, D., Kemmerer, D., Adolphs, R., Damasio, H., & Damasio, A. R. (2003). Neural correlates of conceptual knowledge for actions. *Cognitive Neuropsychology*, *20*(3-6), 409–432.

TURING, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *59*(236), 433–460.

Vacharkulksemsuk, T., & Fredrickson, B. (2012). Strangers in sync: Achieving embodied rapport through shared movements. *Journal of experimental social psychology*, *48*, 399–402.

Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology - J EXP SOC PSYCHOL*, *46*, 693–695.

Van Baaren, R. B., Decety, J., Dijksterhuis, A., van der Leij, A., & van Leeuwen, M. L. (2009). Being imitated: Consequences of nonconsciously showing empathy. *The social neuroscience of empathy*, 31–42.

Van Baaren, R. B., Horgan, T. G., Chartrand, T. L., & Dijkmans, M. (2004). The forest, the trees, and the chameleon: Context dependence and mimicry. *Journal of personality and social psychology*, *86*(3), 453.

Van Baaren, R. B., Maddux, W. W., Chartrand, T. L., De Bouter, C., & Van Knippenberg, A. (2003). It takes two to mimic: Behavioral consequences of self-construals. *Journal of personality and social psychology*, *84*(5), 1093.

Van Baaren, R., Janssen, L., Chartrand, T. L., & Dijksterhuis, A. (2009). Where is the love? the social aspects of mimicry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2381–2389.

van Baaren, R. B., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, *15*(1), 71–74. PMID: 14717835.

van Baaren, R. B., Fockenberg, D. A., Holland, R. W., Janssen, L., & van Knippenberg, A. (2006). The moody chameleon: The effect of mood on non–conscious mimicry. *Social Cognition*, *24*(4), 426–437.

van Leeuwen, M. L., van Baaren, R. B., Martin, D., Dijksterhuis, A., & Bekkering, H. (2009). Executive functioning and imitation: Increasing working memory load facilitates behavioural imitation. *Neuropsychologia*, *47*(14), 3265–3270.

van Leeuwen, M. L., Veling, H., van Baaren, R. B., & Dijksterhuis, A. (2009). The influence of facial attractiveness on imitation. *Journal of Experimental Social Psychology*, *45*(6), 1295–1298.

van Swol, L. M. (2003). The effects of nonverbal mirroring on perceived persuasiveness, agreement with an imitator, and reciprocity in a group discussion. *Communication Research*, *30*(4), 461–480.

van Ulzen, N., Lamoth, C., Daffertshofer, A., Semin, G., & Beek, P. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience letters*, *432*, 88–93.

Varlet, M., Marin, L., Lagarde, J., & Bardy, B. (2011). Social postural coordination. *Journal of experimental psychology. Human perception and performance*, *37*, 473–83.

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2015). Trusting a Virtual Driver That Looks, Acts, and Thinks Like You. *Human Factors*, *57*(5), 895–909.

Verberne, F. M. F., Ham, J., Ponnada, A., & Midden, C. J. H. (2013). Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust. In S. Berkovsky & J. Freyne (Eds.), *Persuasive technology* (pp. 234–245). Berlin, Heidelberg: Springer Berlin Heidelberg.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological review*, *20*(2), 158.

Webb, J. T. (1969). Subject speech rates as a function of interviewer behaviour. *Language and Speech*, *12*(1), 54–67. PMID: 5789296.

Webber Jr, C. L., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, 26–94.

Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, *9*(1), 36–45.

Wheeler, L. (1966). Toward a theory of behavioral contagion. *Psychological Review*, *73*(2), 179.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, 1832.

Wilkes-Gibbs, D. (1986). Collaborative processes of language use in conversation. *doctoral dissertation. Stanford University.*

Wiltermuth, S., & Heath, C. (2009). Synchrony and cooperation. *Psychological science*, *20*, 1–5.

Włodarczak, M., Buschmeier, H., Malisz, Z., Kopp, S., & Wagner, P. (2012). Listener head gestures and verbal feedback expressions in a distraction task. In *Feedback behaviors in dialog.*

Wong, S. P.-H., Hartley, N., & Tombs, A. (2017). Behavioural effects of nonconscious mimicry and social intentions. *Australasian Marketing Journal (AMJ)*, *25*(1), 26–37.

Wyatt, T. D. et al. (2003). *Pheromones and animal behaviour: Communication by smell and taste.* Cambridge university press.

Yabar, Y., Johnston, L., Miles, L., & Peace, V. (2006). Implicit behavioral mimicry: Investigating the impact of group membership. *Journal of Nonverbal Behavior*, *30*(3), 97–113.

Yee, N., & Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33*(3), 271–290.

Yngve, V. H. (1970). On getting a word in edgewise. In *Chicago linguistics society, 6th meeting, 1970* (pp. 567–578).

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*(3681), 269–274.

Zhong, C.-B., & Leonardelli, G. J. (2008). Cold and lonely: Does social exclusion literally feel cold? *Psychological Science*, *19*(9), 838–842. PMID: 18947346.

Zibrek, K., Martin, S., & McDonnell, R. (2019). Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Trans. Appl. Percept.* *16*(3), 14:1–14:19.

Zorzi, M., & Umiltá, C. (1995). A computational model of the simon effect. *Psychological Research*, *58*(3), 193–205.

Appendices

Appendix A

Training message

We expect you to attend all lectures and labs and any other events that are part of your modules. If you are absent from College for more than a day or two you will need to inform the Student Support Officer at the earliest opportunity. You are not permitted to be absent for any other reason unless you have prior permission from the Senior Tutor. If something serious such as an illness prevents you from attending, you should report this to the Student Support Officer at your earliest convenience.

Your attendance during your lab sessions will be monitored via barcode scanners. Poor attendance will result in the Senior Tutor/Student Support Officer sending a notice to your Queen Mary email address. If you do not reply to this email within 7 days a record of your poor attendance will be put on your file. This information may be passed to your local authority.

Appendix B

The Persuasive Message of Student ID Card

This is your student ID card, which is also a library card. The ID card is valid for the entire duration of study and serves a number of functions and is required for identification at the University e.g. when collecting your official letters and/or cheques, when sitting examinations and for access to University facilities such as the Sports facilities. All students are required to carry the card when on University premises.

Here is a few things you should remember:

1. The student ID card is the property of the University
2. All students are required to produce their student ID card when accessing certain University facilities
3. A student may be asked to present their ID card as proof of identity by security or any other University staff whilst on University premises
4. The student should not allow anyone else to use their student ID card or disciplinary procedures may be imposed.
5. The student ID card will act as the Library card and users of the Library must comply with Library Regulations.
6. ID cards are mandatory for all University examinations and failure to show a valid ID card could result in the student not being able to sit their exam.

You must be in possession of a valid ID card for the duration of your study, if you leaves the University you must return the card to Student Services. If your ID card is lost or stolen it is the responsibility of you to notify Student Services and the Library. Lost cards are frequently handed in to Student Services and we will return this to the card holder. In order to cover administration costs a fee of £10.00 is charged to replace lost or damaged cards. You will be asked to pay at Financial Services then bring their receipt of payment to Student Services the ID card will then be produced on the same day. You should bring with them another form of identification (bank card, driving license or passport) in order to collect the new ID card. In cases where the ID card has been stolen, the charge is waived providing you has reported the theft to the Police and has produces a crime reference number from the Police at the time of requesting a new ID card.

Appendix C

Questionnaires

Thank you for participate the experiment. Could you please take a few minutes to answer the questionnaire below?

C.1 Demography

1. What's your occupation/subject?
2. What's your nationality?
3. Your age?
4. Your gender? Male/Female

C.2 Agreement

1. I agree with the plan to implement ID cards.
   (1 Strongly disagree - 7 Strongly agree)
2. I think the proposed ID cards are valuable.
   (1 Strongly disagree - 7 Strongly agree)
3. I think the proposed ID cards are workable.
   (1 Strongly disagree - 7 Strongly agree)
4. I think the proposed ID cards are needed.
   (1 Strongly disagree - 7 Strongly agree)

C.3 Social Presence

1. The presenter was friendly.

   (1 Strongly disagree - 7 Strongly agree)

2. The presenter was likeable.

   (1 Strongly disagree - 7 Strongly agree)

3. The presenter was honest.

   (1 Strongly disagree - 7 Strongly agree)

4. The presenter was competent.

   (1 Strongly disagree - 7 Strongly agree)

5. The presenter was warm.

   (1 Strongly disagree - 7 Strongly agree)

6. The presenter was informed.

   (1 Strongly disagree - 7 Strongly agree)

7. The presenter was credible.

   (1 Strongly disagree - 7 Strongly agree)

8. The presenter was modest.

   (1 Strongly disagree - 7 Strongly agree)

9. The presenter was approachable.

   (1 Strongly disagree - 7 Strongly agree)

10. The presenter was interesting.

    (1 Strongly disagree - 7 Strongly agree)

11. The presenter was trustworthy.

    (1 Strongly disagree - 7 Strongly agree)

12. The presenter was sincere.

    (1 Strongly disagree - 7 Strongly agree)

C.4 Impressive

1. To what extent have you enjoyed the experience just now?

   (1 Not at all - 7 Very much)

2. To what extent do you want to meet him/her again in current situation?

   (1 Not at all - 7 Very much)

3. To what extent do you feel him/her is isolated?

   (1 Not at all - 7 Very much)

4. Would you like to meet him/her again?

   (1 Not at all - 7 Very much)

5. To what extent you felt comfortable with him/her?

   (1 Not at all - 7 Very much)

6. How cooperative were him/her?

   (1 Not at all - 7 Very much)

7. Did him/her make you feel self-conscious or embarrassed?

   (1 Not at all - 7 Very much)

C.5 Realism

1. Please list any thoughts you may have about the interaction with the presenter.
2. Was there anything unusual about this interaction?
3. Please write a few sentences about the presenter's LIP movements while speaking.
4. Please write a few sentences about the presenter's HEAD movements while speaking.

C.6 Overall

Could you please write a few words about the study (your feeling, suggestions)?

Pro forma information sheet and consent form

Queen Mary
University of London

Information sheet

Research study [Experimenting with Transformed Social Interaction in Immersive Virtual Environment]: information for participants

We would like to invite you to be part of this research project, if you would like to. You should only agree to take part if you want to, it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please ask if there is anything that is not clear or if you would like more information.

If you decide to take part you will be asked to sign the attached form to say that you agree.

You are still free to withdraw at any time and without giving a reason.

This is one of a series of studies to understand how we could communicate with each other effectively within immersive virtual environment (IVE) or virtual reality (VR). The whole process will take about 30 minutes.

During the experiment, you will wearing a black suit with marks in order to track your movement and rendering into virtual environment. You will wear a Head Mounted Device (HMD) and headphones which will present the virtual environment to you.

After you having immersed into the virtual environment, you will find yourself in an office. You can either look or walking around to get used to the environment. You will then meet a virtual character who will greet you, and then the experimental trials will start when you are ready. The virtual character will present some messages to you in the immersive virtual environment, face to face.

After that, you will need to finish an online survey. Please note you will be video taped during the experiment. And your captured movement will be stored only for research use. You are free to stop the experiment at any time without give any reason.

Thank you for your participation. Please do not discuss this study with others for about one month, since the study is continuing.

IMPORTANT
When people use virtual reality systems, some people sometimes experience some

degree of nausea. If at any time you wish to stop taking part in the study due to this or any other reason, please just say so and we will stop.

There has been some research, which suggests that people using head-mounted displays might experience some disturbances in vision afterwards. No long term studies are known to us, but the studies which have been carried out do testing after about 30 minutes, and find the effect is still sometimes there.

There have been various reported side effects of using virtual reality equipment, such as 'flashbacks'.

With any type of video equipment there is a possibility that an epileptic episode may be generated. This, for example, has been reported for computer video games.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

If you have any questions or concerns about the manner in which the study was conducted please, in the first instance, contact the researcher responsible for the study. If this is unsuccessful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W104, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Queen Mary University of London

Consent form

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: Experimenting with Transformed Social Interaction in Immersive Virtual Environment
Queen Mary Ethics of Research Committee Ref: 1552

- Thank you for considering taking part in this research. The person organizing the research must explain the project to you before you agree to take part.
- If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.
- *I understand that if I decide at any other time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately.*
- *I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.*

Participant's Statement:

I ____ agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and the Information Sheet about the project, and understand what the research study involves.

Signed: Date:

Investigator's Statement:

I ____ confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the volunteer

Pro forma information sheet and consent form

Queen Mary
University of London

Information sheet

Research study [Experimenting with Transformed Social Interaction in Immersive Virtual Environment]: information for participants

We would like to invite you to be part of this research project, if you would like to. You should only agree to take part if you want to, it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please ask if there is anything that is not clear or if you would like more information.

If you decide to take part you will be asked to sign the attached form to say that you agree.

You are still free to withdraw at any time and without giving a reason.

This is one of a series of studies to understand how we could run student information service effectively within immersive virtual environment (IVE) or virtual reality (VR). The whole process will take about 60 minutes.

During the experiment, you will wearing a black suit with marks in order to track your movement and rendering into virtual environment. You will wear a Head Mounted Device (HMD) and headphones which will present the virtual environment to you.

After you having immersed into the virtual environment, you will find yourself in an office. You can either look or walking around to get used to the environment. You will have some time get yourself comfortable. If you are ready, you will then meet a virtual character who will greet you, and then the experimental trial will start when you are ready. The virtual character will present some messages to you in the immersive virtual environment, face to face. At the second trial, you will be asked to read a message in front of a virtual character.

After that, you will need to finish an online survey. Please note you will be video taped during the experiment. And your captured movement will be stored only for research use. You are free to stop the experiment at any time without give any reason.

Thank you for your participation. Please do not discuss this study with others for about one month, since the study is continuing.

IMPORTANT

When people use virtual reality systems, some people sometimes experience some degree of nausea. If at any time you wish to stop taking part in the study due to this or any other reason, please just say so and we will stop.

There has been some research, which suggests that people using head-mounted displays might experience some disturbances in vision afterwards. No long term studies are known to us, but the studies which have been carried out do testing after about 30 minutes, and find the effect is still sometimes there.

There have been various reported side effects of using virtual reality equipment, such as 'flashbacks'.

With any type of video equipment there is a possibility that an epileptic episode may be generated. This, for example, has been reported for computer video games.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

If you have any questions or concerns about the manner in which the study was conducted please, in the first instance, contact the researcher responsible for the study. If this is unsuccessful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W104, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Queen Mary
University of London

Consent form

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: Experimenting with Transformed Social Interaction in Immersive Virtual Environment
Queen Mary Ethics of Research Committee Ref: ____1780_______

- Thank you for considering taking part in this research. The person organizing the research must explain the project to you before you agree to take part.
- If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.
- *I understand that if I decide at any other time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately.*
- *I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.*

Participant's Statement:

I ____ agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and the Information Sheet about the project, and understand what the research study involves.

Signed: Date:

Investigator's Statement:

I ____ confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the volunteer